

Systems biology

SBML-PET: a Systems Biology Markup Language-based parameter estimation tool

Zhike Zi and Edda Klipp*

Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany

Received on June 15, 2006; revised on August 2, 2006; accepted on August 14, 2006

Advance Access publication August 22, 2006

Associate Editor: Satoru Miyano

ABSTRACT

Summary: The estimation of model parameters from experimental data remains a bottleneck for a major breakthrough in systems biology. We present a Systems Biology Markup Language (SBML) based Parameter Estimation Tool (SBML-PET). The tool is designed to enable parameter estimation for biological models including signaling pathways, gene regulation networks and metabolic pathways. SBML-PET supports import and export of the models in the SBML format. It can estimate the parameters by fitting a variety of experimental data from different experimental conditions. SBML-PET has a unique feature of supporting event definition in the SBML model. SBML models can also be simulated in SBML-PET. Stochastic Ranking Evolution Strategy (SRES) is incorporated in SBML-PET for parameter estimation jobs. A classic ODE Solver called ODEPACK is used to solve the Ordinary Differential Equation (ODE) system.

Availability: <http://sysbio.molgen.mpg.de/SBML-PET/>. The website also contains detailed documentation for SBML-PET.

Contact: klipp@molgen.mpg.de, zhike_zi@molgen.mpg.de

1 INTRODUCTION

The aim of parameter estimation for biochemical network models is to find the most feasible parameters that reproduce these experimental results according to a given set of experimental data. Several efforts have been dedicated to this problem. Some researchers focus on the development of optimization methods (Tsai and Wang, 2005; Klipp and Heinrich, 1994; Mendes and Kell, 1998; Kremling *et al.*, 2004; Zak *et al.*, 2003; Matsubara *et al.*, 2006; Runarsson and Yao, 2000; Gadkar *et al.*, 2005), others pay attention to making the methods available for the users. For example, Grid Cellware (Dhar *et al.*, 2005) and OBIYagns (Kimura *et al.*, 2004) are grid-based tools with parameter estimator. Gepasi (Mendes, 1993) and Copasi (<http://www.copasi.org>), both complex pathway simulators, enable parameter estimation with different optimization methods using data from time course or steady state experiments. The C library libSRES (Ji and Xu, 2006) is composed of a parameter estimator mainly implementing stochastic ranking evolution strategy.

Although there are some simulation programs with parameter estimation facility, it is still hard for systems biologists to apply them to their specific problems. First, none of the currently existing software supports models in which events have been defined, although events are very common and important for biological

experiments. Most experimental data refer to a specific event in the experiment like changing signal source or strength at a specific time, or blocking protein synthesis during the experiment. Currently available software is difficult to apply to such cases. Second, sometimes it is impossible to directly measure the concentration of the components in the biological model. Instead, only the sum amount of several components or the relative change of a component can be measured. Parameter estimators should support any kind of experimental data, e.g. measurement error in the experimental data, normalized data or any mathematical expression with the concentration of the components.

In order to satisfy the specific customized requirements for parameter estimation, we developed an SBML-based Parameter Estimation Tool (SBML-PET). SBML-PET is designed to do parameter estimation work for biological models including signaling pathways, gene regulation networks and metabolic pathways. To the best of our knowledge, SBML-PET has a unique feature of supporting the models including events. It also supports a variety of data from different experimental conditions and the data can be expressed with common mathematical expression (e.g. trigonometric and transcendental functions).

2 FEATURES

The following is a summary of the features of SBML-PET:

SBML-PET supports model import and export in SBML format, a widely accepted standard for the exchange of biochemical network models. All estimated parameters will be saved in a new SBML file, which can be imported by other SBML supported simulation softwares.

SBML-PET supports event structures that describe the time and form of explicit, instantaneous, discontinuous state changes in the model.

Frequently, experimental data are (1) obtained from different labs and (2) produced under different conditions. SBML-PET can estimate the parameters using all such diverse types of data.

Concentration measurements are often indirect. SBML-PET supports any kind of experimental data, e.g. normalized data or common mathematical expression involving the concentration of components.

SBML-PET supports common mathematical expressions for the qualitative and quantitative description of the model, such as constraints for parameters or combinations thereof.

SBML-PET also supports the standard deviation for the data and the noise or measurement error existed in the experiments.

*To whom correspondence should be addressed.

3 METHODS

SBML-PET incorporates Stochastic Ranking Evolution Strategy (SRES) for parameter estimation jobs. SRES is a (μ, λ) -ES evolutionary optimization algorithm that uses stochastic ranking as the constraint handling technique (Runarsson and Yao, 2000). Moles *et al.* (2003) compared several common global optimization methods and showed that SRES achieves the best result in the case of a three-step pathway. The performance of SRES has been extensively tested on various pathway parameter estimation problems and is found to be satisfactory.

All solutions of ODE systems in SBML-PET are computed by ODEPACK (Hindmarsh, 1983), which is a collection of FORTRAN ODE solvers. For models without events, SBML-PET uses LSODA, which solves ODE systems with automatic method selection between non-stiff (Adams) and stiff (Backward Differentiation Formula, BDF) methods. For models including events, LSODAR is chosen. LSODAR is a variant of LSODA with a root finding capability added. LSODARs root finder is used in this application to locate events defined in the model.

4 IMPLEMENTATION

SBML-PET runs in a command interactive mode environment. It is tested both in Linux and Cygwin on Windows. It is easy to install and to use. Following the guide information printed on the screen, you can easily finish the parameter estimation work for your SBML model.

Before implementing SBML-PET, the users need to prepare the data file following the instructions for the data file. Detailed information about the preparation of data file is described in the manual document.

Five examples covering different cases of parameter estimation are provided in the manual document. The simple parameter estimation for the model of Michaelis-Menten equations clearly shows how to use SBML-PET. A yeast aging model and a JAK-STAT model (Swameye *et al.*, 2003) demonstrate how SBML-PET supports events. Real experimental data with mathematical expression are used to estimate parameters in JAK-STAT model and the model of irreversible inhibition of HIV proteinase (Petr 1996; Mendes and Kell, 1998). Parameter estimation for different experimental conditions (different extra-cellular tryptophan concentrations) are shown in the dynamic model of *Escherichia coli* tryptophan operon (Sharad *et al.*, 2003).

The time to find the best fit solution depends on the complexity of the ODE system (the number of ODEs and parameters), the number of experimental data and the CPU speed of the computer. For simple problem (with a few number of species and data), it takes minutes and hours to get the best result. For complex problem (with a lot of species or data), it takes ~ 1 day to get the best result. The time of finding best solution for the tested examples shown in manual file is

~ 15 min for model of Michaelis-Menten equations, Yeast Aging model and JAK-STAT model, 5 h for HIV Model and 2 h for dynamic model of *E.coli* tryptophan operon. The examples are run in Windows 2000 with CPU of AMD Athlon 64 3200, 2.2 GHz.

ACKNOWLEDGEMENTS

This work was supported by an IMPRS fellowship to Z.Z. and by the Berlin Center of Genome-based Bioinformatics founded by the German Ministry of Education and Research.

Conflict of Interest: none declared.

REFERENCES

- Hindmarsh,A.C. (1983) ODEPACK, a systematized collection of ODE solvers. In Stepleman,R.S., Carver,M., Peskin,R., Ames,W.F. and Vichnevetsky,R. (eds), *Scientific Computing*, Amsterdam, North-Holland Publishing Co., 55–64.
- Dhar,P.K. *et al.* (2005) Grid cellware: the first grid-enabled tool for modelling and simulating cellular processes. *Bioinformatics*, **21**, 1284–1287.
- Gadkar,K.G. *et al.* (2005) Iterative approach to model identification of biological networks. *BMC Bioinformatics*, **6**, 155.
- Ji,X. and Xu,Y. (2006) libSRES: a C library for stochastic ranking evolution strategy for parameter estimation. *Bioinformatics*, **22**, 124–126.
- Kimura,S. *et al.* (2004) OBIYagns: a grid-based biochemical simulator with a parameter estimator. *Bioinformatics*, **20**, 1646–1648.
- Klipp,E. and Heinrich,R. (1994) Evolutionary optimization of enzyme kinetic parameters; effect of constraints. *J. Theor. Biol.*, **171**, 309–323.
- Kremling,A. *et al.* (2004) A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Res.*, **14**, 1773–1785.
- Matsubara,Y. *et al.* (2006) Parameter estimation for stiff equations of biosystems using radial basis function networks. *BMC Bioinformatics*, **7**, 230.
- Mendes,P. (1993) Gepasi—a software package for modeling the dynamics, steady-states and control of biochemical and other systems. *Comput. Appl. Biosci.*, **9**, 563–571.
- Mendes,P. and Kell,D. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, **14**, 869–883.
- Moles,C.G. *et al.* (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, **13**, 2467–2474.
- Petr,K. (1996) Program dynafit for the analysis of enzyme kinetic data: Application to hiv proteinase. *Analytical Biochemistry*, **237**, 260–273.
- Runarsson,T.P. and Yao,X. (2000) Stochastic ranking for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.*, **4**, 284–294.
- Sharad,B. *et al.* (2003) Dynamic model of *Escherichia coli* tryptophan operon shows an optimal structural design. *Eur. J. Biochem.*, **270**, 2644–2651.
- Swameye,I. *et al.* (2003) Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc. Natl Acad. Sci. USA*, **100**, 1028–1033.
- Tsai,K.Y. and Wang,F.S. (2005) Evolutionary optimization with data collocation for reverse engineering of biological networks. *Bioinformatics*, **21**, 1180–1188.
- Zak,D.E. *et al.* (2003) Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of an *in silico* network. *Genome Res.*, **13**, 2396–2405.