Bioinformatics Advance Access published June 14, 2015

Systems biology

REDEMPTION: Reduced Dimension Ensemble Modeling and Parameter Estimation

Yang Liu[1,2], Erica Manesso[1,2], and Rudiyanto Gunawan[1,2,*]

[1]Institute for Chemical and Bioengineering, ETH Zurich, 8093 Zurich, Switzerland, [2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Associate Editor: Dr. Jonathan Wren

ABSTRACT

Summary: Here, we present REDEMPTION (**Re**duced **D**imension **E**nsemble **M**odeling and **P**arameter estima**tion**), a toolbox for parameter estimation and ensemble modeling of ordinary differential equations (ODEs) using time-series data. For models with more reactions than measured species, a common scenario in biological modeling, the parameter estimation is formulated as a nested optimization problem based on incremental parameter estimation strategy. REDEMPTION also includes a tool for the identification of an ensemble of parameter combinations that provide satisfactory goodness-of-fit to the data. The functionalities of REDEMPTION are accessible through a MATLAB user interface (UI), as well as through programming script. For computational speed-up, REDEMPTION provides a numerical parallelization option using MATLAB Parallel Computing toolbox.

Availability: REDEMPTION can be downloaded from http://www.cabsel.ethz.ch/tools/redemption.

Contact: rudi.gunawan@chem.ethz.ch

1 INTRODUCTION

Mathematical modeling plays an increasingly important role in understanding and predicting biological phenomena. In this regard, ODE models are often created to describe the dynamic behavior of biological systems based on mass or molar conservation, as follow:

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{S}\mathbf{v}(\mathbf{X}(t),\mathbf{p}); \quad \mathbf{X}(0) = \mathbf{X}_0, \tag{1}$$

where t denotes time, $\mathbf{X}(t)$ is the species concentration vector, $\mathbf{S}$ is the stoichiometric or connectivity matrix, $\mathbf{v}(\mathbf{X}(t),\mathbf{p})$ is the reaction rate or flux equation vector, $\mathbf{p}$ denotes the kinetic parameter vector, and $\mathbf{X}_0$ denotes the initial concentration vector. The estimation of unknown kinetic parameters from time-series data is the most challenging step in the creation of such biological models, due to: (1) high computational cost associated with global optimizations and with integrating ODEs (Chou and Voit, 2009), and (2) the lack of parameter identifiability (Srinath and Gunawan, 2010).

The high computational complexity could be addressed by performing the parameter estimation incrementally. Here, time-series data are first smoothened and differentiated, and subsequently the dynamic flux values are computed by viewing Eq. (1) as an algebraic equation. The parameters are estimated one flux at a time while avoiding ODE integrations (Voit and Almeida, 2004). However, such an approach requires that $\mathbf{S}$ has a full column rank, which is often not satisfied in the modeling of biological systems.

We recently developed incremental parameter estimation (IPE) and integrated flux parameter estimation (IFPE) to address the limitation related to the matrix $\mathbf{S}$. In IPE (IFPE), we formulated the parameter estimation as a nested optimization, where the outer optimization was performed over a subset of parameters corresponding to the independent (integrated) fluxes, defined such that given their values, the remaining (integrated) fluxes could be estimated from the slope of time-series data (Jia *et al.*, 2012a, Liu and Gunawan, 2014). The inner optimizations involved estimating parameters associated with the dependent flux subset using their estimated (integrated) flux values, performed one flux at a time.

Meanwhile, incomplete parameter identifiability implies that many parameter combinations could fit the data equally well (Jia, et al., 2012b). In ensemble modeling, instead of generating one best-fit parameter estimate, one looks for a set of parameter combinations which fit the data within an acceptable prediction accuracy, e.g. an upper bound for the error function. We extended our incremental estimation strategy above such that the identification of parameter ensemble could be efficiently performed over smaller parameter dimensions (Jia et al., 2012b).

REDEMPTION provides a UI for efficient parameter estimation and ensemble modeling based on IPE and IFPE in MATLAB. The functions within REDEMPTION were designed to address a common scenario in biological modeling, where the number of reactions exceeds that of the (measured) species. When this scenario does not apply, REDEMPTION automatically reverts to a standard incremental parameter estimation. For ensemble modeling, REDEMPTION employs an efficient parameter exploration algorithm HYPERSPACE, which combines an Adaptive Monte Carlo method and a multiple ellipsoids based sampling (Zamora-Sillero, et al., 2011).

2 MAIN FEATURES

Model and data specifications: REDEMPTION's UI starts with the Main window (see Fig. 1(**a**)), from which users can access all functionalities. The

*To whom correspondence should be addressed.

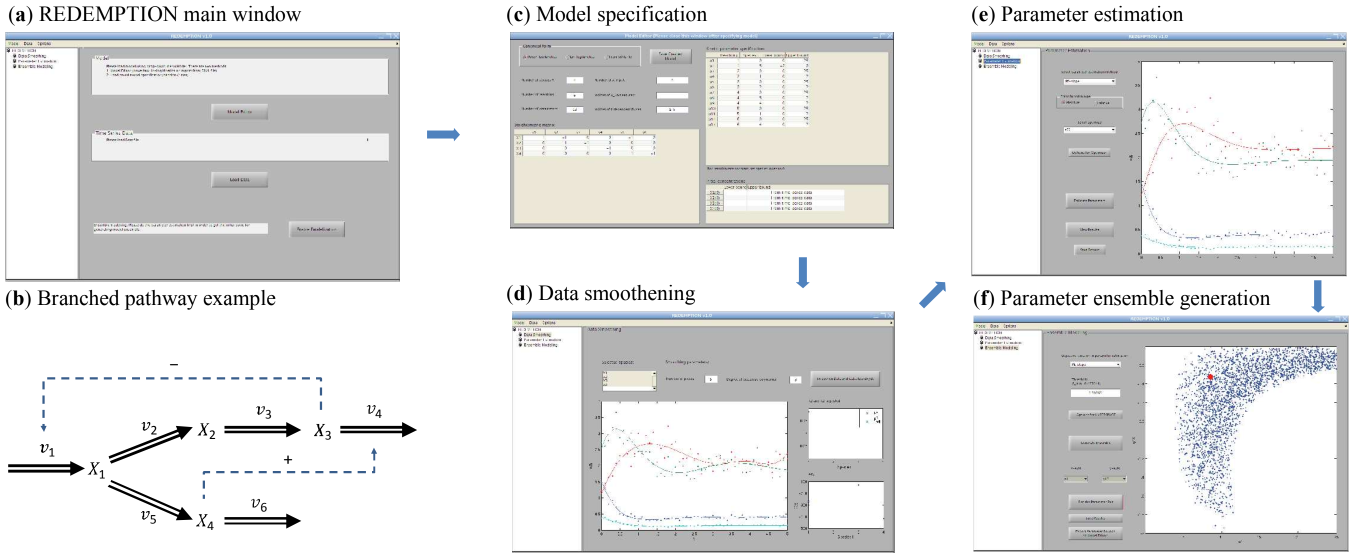

Fig. 1. Workflow of Parameter Estimation and Ensemble Generation in REDEMPTION applied to Branched Pathway Example.

ODE model equations can be specified manually through the Model Editor using power-law or linear-logarithmic (lin-log) kinetics (see Fig. 1(**c**)), or by importing an SBML file. In addition, REDEMPTION requires upper and lower bound values for the model parameters. A parameter will be estimated from data when the upper and lower bounds differ. Users also need to provide the time-series concentration data in Comma-Separated Values (CSV) format. For data pre-processing, REDEMPTION includes piecewise polynomial spline-fitting, where users can adjust the number of pieces and the order of polynomials (see Fig 1(**d**)).

Parameter estimation: Fig. 1(**e**) depicts REDEMPTION's parameter estimation UI. For parameter estimation, REDEMPTION provides four variants of the IPE and IFPE methods and two error functions using sum of squares of the absolute or relative model prediction errors (see User Guide for more details). Here, users can select MATLAB's Genetic Algorithm (GA) or enhanced Scatter Search (eSS) (Egea, et al., 2010) as the global optimization algorithm.

Ensemble modeling: After a successful run of parameter estimation, users can generate parameter ensemble corresponding to parameter combinations whose error function values are smaller than a user-specified threshold (see Fig. 1(**f**)). The optimal error function value obtained from the prior parameter estimation step is provided as a reference for setting this threshold. The appropriate threshold could for example be calculated by bootstrapping the original data (Jia, et al., 2012b). Fig. 1(**f**) shows a 2D parameter projection of the generated parameter ensemble, in which each blue dot represents a parameter combination in the ensemble and the red dot corresponds to the optimal solution of the parameter estimation step. To include an additional dataset, users can perform parameter estimation and ensemble modeling again using the additional data and new parameter bounds based on the parameter ensemble (see User Guide for more information).

3 EXAMPLES

Fig. 1 illustrates the workflow of REDEMPTION for parameter estimation and ensemble generation using a branched pathway example as shown in Fig. 1**(b)** (Voit and Almeida, 2004). REDEMPTION also includes the lin-log modeling of *Lactococcus lactis* glycolytic pathway as another example (del Rosario, et al., 2008). Details of these examples can be found in the User Guide.

4 CONCLUSIONS

REDEMPTION provides an integrated, numerically efficient tool for estimating kinetic parameters and parameter ensemble of ODE models. The tools within REDEMPTION are accessible through a user-friendly MATLAB UI and applicable for ODE models with power-law or lin-log kinetics, and those in SBML format.

AUTHORS' CONTRIBUTIONS

YL and RG designed the toolbox. YL programmed the toolbox. YL, EM and RG coordinated the testing and wrote the manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Julio Banga for providing the eSS algorithm, and members of CABSEL for testing the toolbox.

Funding: This work was supported by funding from ETH Zurich and Swiss National Science Foundation (Grant # 137614).

Conflict of Interest: none declared.

REFERENCES

Chou, I.C. and Voit, E.O. (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems, *Math Biosci*, **219**, 57-83.

del Rosario, R.C., Mendoza, E. and Voit, E.O. (2008) Challenges in lin-log modelling of glycolysis in Lactococcus lactis, *IET Syst Biol*, **2**, 136-149.

Egea, J.A., Marti, R. and Banga, J.R. (2010) An evolutionary method for complex-process optimization, *Comput Oper Res*, **37**, 315-324.

Jia, G., Stephanopoulos, G. and Gunawan, R. (2012a) Incremental parameter estimation of kinetic metabolic network models, *BMC Syst Biol*, **6**, 142.

Jia, G., Stephanopoulos, G. and Gunawan, R. (2012b) Ensemble Kinetic Modeling of Metabolic Networks from Dynamic Metabolic Profiles, *Metabolites*, **2**, 891-912.

Liu, Y. and Gunawan, R. (2014) Parameter estimation of dynamic biological network models using integrated fluxes, *BMC Syst Biol*, **8**, 127.

Srinath, S. and Gunawan, R. (2010) Parameter identifiability of power-law biochemical system models, *J Biotechnol*, **149**, 132-140.

Voit, E.O. and Almeida, J. (2004) Decoupling dynamical systems for pathway identification from metabolic profiles, *Bioinformatics*, **20**, 1670-1681.

Zamora-Sillero, E., *et al.* (2011) Efficient characterization of high-dimensional parameter spaces for systems biology, *BMC Syst Biol*, **5**, 142.