BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 1 2014, pages 137–138
doi:10.1093/bioinformatics/btt631

Systems biology

Advance Access publication November 4, 2013

StochDecomp—Matlab package for noise decomposition in stochastic biochemical systems

Tomasz Jetka[1], Agata Charzyńska[2], Anna Gambin[3], Michael P.H. Stumpf[4,*] and Michał Komorowski[1,*]

[1]Institute of Fundamental Technological Research, [2]Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland, [3]Faculty of Mathematics Informatics and Mechanics, Institute of Informatics, University of Warsaw, Warsaw, Poland and [4]Division of Molecular Biosciences, Imperial College London, London, UK

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Stochasticity is an indispensable aspect of biochemical processes at the cellular level. Studies on how the noise enters and propagates in biochemical systems provided us with non-trivial insights into the origins of stochasticity, in total, however, they constitute a patchwork of different theoretical analyses.

Results: Here we present a flexible and widely applicable noise decomposition tool that allows us to calculate contributions of individual reactions to the total variability of a system's output. With the package it is, therefore, possible to quantify how the noise enters and propagates in biochemical systems. We also demonstrate and exemplify using the JAK-STAT signalling pathway that the noise contributions resulting from individual reactions can be inferred from data experimental data along with Bayesian parameter inference. The method is based on the linear noise approximation, which is assumed to provide a reasonable representation of analyzed systems.

Availability and implementation: http://sourceforge.net/p/stochdecomp/

Contact: mkomor@ippt.gov.pl or m.stumpf@imperial.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on August 15, 2013; revised on October 28, 2013; accepted on October 29, 2013

1 INTRODUCTION

The question which molecular species or parts of a network contribute most of the variability of a system or are responsible for most of the information loss has attracted much attention in recent years. Numerous studies have analyzed noise in signalling networks in detail and decomposed the noise into contributions attributable to fluctuations in messenger RNA and protein. Current software implementations offer a broad range of stochastic modeling methods to analyze stochastic properties of biochemical dynamics (Andrews *et al.*, 2010; Thomas *et al.*, 2012). These tools, however, focus only indirectly on origins and propagation of stochasticity. To our knowledge, a software package to provide decomposition of noise into individual sources has been lacking. Recently, we developed (Komorowski *et al.*, 2013) a flexible and simple method to analyze how the structure of biochemical networks gives rise to noise in its outputs. In principle, this allows us to efficiently calculate the contribution each reaction makes to the variability in all concentrations for any network, which can be modelled within the linear noise approximation (LNA) framework. Origins of variability can be therefore assigned to individual reactions and arbitrarily defined network components.

Moreover, if experimental data are available and a posterior of model parameters can be generated, the contribution of individual reactions can be estimated from data along with the parameters. Later in the text we provide a general description of the package. Details are presented in the Supplementary Information, which includes theoretical foundations of the method, user manual and examples. In a comprehensive analysis of the JAK-STAT signalling pathway, we infer individual contributions from experimental data published in Swameye *et al.* (2003).

2 METHODS

The LNA was used to model stochastic chemical kinetics (van Kampen, 2007, Komorowski 2009). In the LNA the covariance Σ, a matrix quantifying the noise in every network component is represented in form of the deterministic ordinary differential equations (ODEs) (see Supplementary Material for details)

$$\frac{d\Sigma}{dt} = A(t)\Sigma + \Sigma A(t)^T + D(t) \quad (1)$$

Because (1) is linear in Σ, and D decomposes into a sum across reactions, Σ likewise decomposes into a sum across reactions (Komorowski *et al.*, 2013)

$$\Sigma = \Sigma^{(1)} + \ldots + \Sigma^{(r)} \quad (2)$$

where r denotes the number of reactions in the system. From a specification of the network, we calculate the response matrix A, which describes how the network state instantaneously responds to fluctuations, and the dissipation matrix D, which describes the contribution of count noise. This enables us to identify the origins of cell-to-cell variability in dynamical biochemical systems and pinpoint, if warranted, individual reactions.

2.1 Implementation

The package is implemented as a set of Matlab functions. To be analyzed, model needs to be defined in terms of a stoichiometry matrix, a Matlab function containing reaction rates and a vector of parameter values.

*To whom correspondence should be addressed

A function that generates this definition files from an Systems Biology Markup Language (SBML) file is provided. The definition files are used to generated a set of ODEs using the Matlab symbolic toolbox. Equations are then solved using the Matlab ODE solver, and solutions provide the variance decomposition. Functions to providing graphical output are also implemented.

2.2 Applicability

The package assumes that the LNA is a reasonable approximation of modelled systems. Generally, this is the case if the number of each of the interacting reacting molecules is large and the system is monostable. Detailed discussions on the validity of the LNA are presented in Ramaswamy *et al.* (2012) and Wallace *et al.* (2012). Accounting for this limitation, the tool allows us to take any modelled network and efficiently calculate the contribution each reaction makes to the variability in all concentrations, specifically (i) symbolically generate ODEs describing the system and individual reaction contributions, (ii) numerically compute variance decomposition and visualize obtained results and (iii) infer contributions of individual reactions from experimental data if posterior distribution is provided. The flow chart describing input–output relationship of the package is presented in Figure 1.

2.3 Biological relevance

Using experimental data of Swameye *et al.* (2003), we inferred the sources of variability in the JAK-STAT signalling pathway. First, we used the Prediction Uncertainty Analysis (PUA) Matlab package (Vanlier *et al.*, 2012) to generate posterior distribution of model parameters as described in Vanlier *et al.* (2012). Second, the parameter posterior was translated by our package into a posterior of noise contributions. The tool revealed the following insight about variability in the nuclear concentration of STAT complexes, which is a factor activating a downstream response: (i) in the absence of extrinsic noise, the fluctuations in the number of nuclear complexes originate largely from trafficking of the complexes into the nucleus. (ii) In the presence of the extrinsic noise, understood as fluctuations in Epo concentration, the network acts as a low pass filter. The extrinsic noise is major source of variability if the fluctuations in Epo concentration are slow. (iii) The overall variability of the nuclear concentration of STAT complexes is relatively insensitive to parameters. Contributions of certain reactions, however, are sensitive and change by an order of magnitudes for the parameters within the posterior (see Supplementary Material for details).

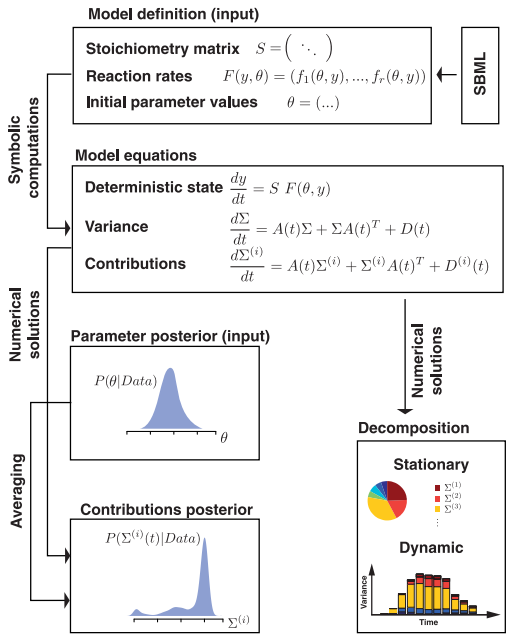

Fig. 1. Input-output flow chart of the StochDecomp package. The stoichiometry matrix, reaction rates and initial parameter guesses, which define the stochastic model, are translated by means of the symbolic computations into the set of deterministic ODEs (see Supplementary Material for details). Conversion from SBML is also implemented. Stored ODEs are then used to numerically analyze the stochastic model, particularly decompose the variability into contributions resulting from each of the model reactions either in steady-state or out-of-steady-state. Though not explicitly stated, the covariance matrix Σ and contributions $\Sigma^{(i)}$ depend on the parameter vector θ. Therefore, if experimental data together with parameter posterior are available, individual contributions can be inferred along with the parameters

3 DISCUSSION

StochDecomp is a novel computationally efficient and integrative Matlab package for computational analysis of noise origin in the stochastic models of biochemical reactions. The ability to dissect noise propagation through biological systems does enable to better understand the role of noise in function and evolution, and will also help synthetic biologists to either harness or dampen the effects of noise in molecular signalling and response networks.

Funding: Foundation for Polish Science (HOMING 2011-3/4) (to T.J. and M.K.); Research fellowship (POKL.04.01.01-00-051/10-00) (to A.C.); National Science Center (2011/01/B/NZ2/00864) and the Biocentrum-Ochota project (POIG 02.03.00-00-003/09) (A.G.); and BBSRC (BB/G020434/1) (to M.P.H.S.). M.P.H.S. is a Royal Society Wolfson Research Merit Award holder. M.K. is EMBO Installation Grantee.

Conflict of Interest: none declared.

REFERENCES

Andrews,S.S. *et al.* (2010) Detailed simulations of cell biology with smoldyn 2.1. *PLoS Comput. Biol.*, **6**, e1000705.

Komorowski,M. *et al.* (2009) Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, **10**, 343.

Komorowski,M. *et al.* (2013) Decomposing noise in biochemical signalling systems highlights the role of protein degradation. *Biophys. J.*, **104**, 1783–1793.

Ramaswamy,R. *et al.* (2012) Discreteness-induced concentration inversion in mesoscopic chemical systems. *Nat. Commun.*, **3**, 779.

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.

Thomas,P. *et al.* (2012) Intrinsic noise analyzer: a software package for the exploration of stochastic biochemical kinetics using the system size expansion. *PLoS One*, **7**, e38518.

van Kampen,N.G. (2007) *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library, Elsevier Science.

Vanlier,J. *et al.* (2012) An integrated strategy for prediction uncertainty analysis. *Bioinformatics*, **28**, 1130–1135.

Wallace,E. *et al.* (2012) Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET Syst. Biol.*, **6**, 102–115.