

Systems biology

SELANSI: a toolbox for simulation of stochastic gene regulatory networks

Manuel Pájaro¹, Irene Otero-Muras¹, Carlos Vázquez² and Antonio A. Alonso^{1,*}

¹BioProcess Engineering Group, IIM-CSIC, Spanish National Research Council, 36208 Vigo, Spain and ²Department of Mathematics, University of A Coruña Campus Elviña s/n, 15071 A Coruña, Spain

*To whom correspondence should be addressed.

Associate Editor: Oliver Stegle

Received on May 31, 2017; revised on September 8, 2017; editorial decision on October 7, 2017; accepted on October 10, 2017

Abstract

Motivation: Gene regulation is inherently stochastic. In many applications concerning Systems and Synthetic Biology such as the reverse engineering and the *de novo* design of genetic circuits, stochastic effects (yet potentially crucial) are often neglected due to the high computational cost of stochastic simulations. With advances in these fields there is an increasing need of tools providing accurate approximations of the stochastic dynamics of gene regulatory networks (GRNs) with reduced computational effort.

Results: This work presents SELANSI (SEmi-LAgrangian Simulation of GRNs), a software toolbox for the simulation of stochastic multidimensional gene regulatory networks. SELANSI exploits intrinsic structural properties of gene regulatory networks to accurately approximate the corresponding Chemical Master Equation with a partial integral differential equation that is solved by a semi-lagrangian method with high efficiency. Networks under consideration might involve multiple genes with self and cross regulations, in which genes can be regulated by different transcription factors. Moreover, the validity of the method is not restricted to a particular type of kinetics. The tool offers total flexibility regarding network topology, kinetics and parameterization, as well as simulation options.

Availability and implementation: SELANSI runs under the MATLAB environment, and is available under GPLv3 license at <https://sites.google.com/view/selansi>.

Contact: antonio@iim.csic.es

1 Introduction

Understanding the dynamics of Gene Regulatory Networks (GRNs) requires appropriate modeling and simulation tools to efficiently represent their underlying stochastic nature. At this level, the states of the system are random variables whose evolution can only be predicted probability-wise via the Chemical Master Equation (CME). Unfortunately, such description, which consists of a large dimensional set of coupled ODEs, does not admit a closed form solution in most cases of practical interest (Kryven *et al.*, 2015). Apart from the now classical Stochastic Simulation Algorithm (SSA) (Gillespie, 2007), numerical solutions of the CME range from discrete to continuous approximations. SSA methods including their software

implementations (e.g. Hoops *et al.*, 2006; Maarleveld *et al.*, 2013; Ramsey *et al.*, 2005; Sanft *et al.*, 2011) make use of Monte Carlo type simulations to reconstruct the probability function from a large number of realizations of the stochastic GRN (Gillespie, 2007). Although such methods are generally applicable, they may become prohibitive due to the computational burden.

Discrete approximations such as the finite state projection (FSP) or moment based (MB) methods, reduce the number of state variables by either restricting the states to those most probable, or by computing the most relevant moments of the underlying distribution. Related software tools include StochDynTools by Hespanha *et al.* (2008), CmePy by Hegland *et al.* (2010) (available at <https://>

github.com/hegland/cmepy), MOCA by Schnoerr *et al.* (2015) and SHAVE by Lapin *et al.* (2011). Despite truncation, accuracy typically demands the number of remaining variables in FSP methods to be substantially large, what makes problems computationally involved.

MB methods would be preferable in those situations where one is interested in computing some particularly relevant moments of the distribution rather than its precise form (see Engblom, 2006). However, reconstructing the probability density function (pdf) from the moments is in general, not straight-forward, since this requires assumptions on the form of the pdf (and relates to the accuracy of the moment closure). This might be simple in some cases (e.g. if near Gaussian, two parameters/moments are enough) but not in others, as for instance under nonlinear kinetics or in characterizing multi-modal distributions (Hasenauer *et al.*, 2014; Pajaro *et al.*, 2017).

Finally, Fokker-Plank (FP) based methods are a class of continuous approximations developed on systems size expansions (Hespanha *et al.*, 2008; Thomas *et al.*, 2013), which inherit similar limitations as the MB methods when the resulting distribution diverges from a Gaussian. A collection of software tools that includes FSP and MB methods can be found in the recently developed toolbox CERENA by Kazeroonian *et al.* (2016).

In between discrete and FP approximations, first Friedman *et al.* (2006) for uni-dimensional GRN and most recently Pajaro *et al.* (2017) for multi-dimensional GRN, proposed a partial integral differential equation (PIDE) approximation based on a time-scale separation property, denoted in what follows as protein bursting, which is shared by most GRNs in prokaryotic and eukaryotic organisms (e.g. Dar *et al.*, 2012). The software tool SELANSI (SEmi-LAgrangian Simulation of gene regulatory networks) approximates the CME by the PIDE model in Pajaro *et al.* (2017) and computes its numerical solution by an efficient and scalable semilagrangian method, providing the temporal evolution of the protein probability density function (together with its stationary state).

2 Multidimensional PIDE model

SELANSI uses a generalized description of GRNs in which each protein can interact with its corresponding gene to regulate its own expression (self-regulation) and/or with any other gene(s) in the network (cross-regulation). For n genes, which are expressed into n different protein types, the number of molecules of each protein is encoded in a vector $\mathbf{x} = (x_1, \dots, x_n)$. The PIDE model implemented in SELANSI exploits the protein bursting assumption (i.e. proteins being produced in episodic bursts), which is valid whenever messenger RNA degrades much faster than proteins. As it has been discussed in Pajaro *et al.* (2017) the PIDE model provides good approximations already for degradation rate ratios in the order of 3–5. Under this assumption, the governing equation of the GRN dynamics (Pajaro *et al.*, 2017) reads:

$$\frac{\partial p(t, \mathbf{x})}{\partial t} = \sum_{i=1}^n \left(\frac{\partial}{\partial x_i} [y_x^i x_i p(t, \mathbf{x})] - k_m^i c_i(\mathbf{x}) p(t, \mathbf{x}) + k_m^i \int_0^{x_i} \omega_i(x_i - y_i) c_i(y_i) p(t, y_i) dy_i \right) \quad (1)$$

where $p(t, \mathbf{x})$ denotes the probability distribution function associated to the n proteins expressed in the network. Functions $\omega_i(x_i - y_i)$ describe the conditional probability for proteins

jumping from a state y_i to x_i . Under the bursting condition, these functions follow an exponential-type distribution of the form

$$\omega_i(x_i - y_i) = \frac{1}{b_i} \exp\left(-\frac{x_i - y_i}{b_i}\right) \quad (2)$$

where $b_i = \frac{k_i}{\gamma_m^i}$ denotes translation frequency and relates to burst size.

Network regulation topology as well as the activation/inactivation gene state dynamics are encoded through functions $c_i(\mathbf{x})$. When the promoter switching rates between on and off states are much faster than transcription-translation, the corresponding expressions c_i are defined in terms of Hill-type functions. SELANSI incorporates pre-defined Hill functions but can accommodate other kinetics (for different ratios of switching rates versus transcription-translation) through appropriate c_i expressions.

Equation (1) is defined in a domain $\Omega = \prod_{i=1}^n (0, L_i) \subset \mathbb{R}_+^n$, with $L_i > 0$, and a time interval $[0, T]$. SELANSI approximates the domain Ω by a uniform (protein) mesh and computes the solution on a uniform time grid by a semilagrangian method (Pajaro *et al.*, 2017). The computational burden depends directly on the size of the mesh and indirectly on the number of genes involved (since higher order networks will require more mesh points for equal accuracy). Guidelines about levels of discretization, computational efficiency and accuracy are provided in the user's manual. By way of indication, SELANSI provides affordable computation times for reasonably smooth solutions up to 5 coupled genes.

3 Main features of SELANSI

SELANSI is a toolbox for simulation of stochastic multidimensional gene regulatory networks implemented in Matlab, working on Windows, Linux and MacOS. The SELANSI toolbox offers:

- **High Flexibility:** Gene networks under consideration might involve multiple genes with self and cross regulations, in which genes can be regulated by different transcription factors. The user can specify the size and topology of the network, as well as the kinetics, parameter values, time horizon and discretization levels for simulation.
- **Generality:** The validity of the method is not restricted to a particular type of kinetics (mass action, Hill, etc.). Although input functions of the Hill type are predefined in SELANSI, the user can easily define his/her own input functions by modifying the available templates.
- **High computational efficiency:** The semilagrangian method implemented in SELANSI is proven efficient and scalable. For networks involving 4 to 5 coupled genes, speed-up factors in computation times are typically of two orders of magnitude with respect to SSAs (Pajaro *et al.*, 2017).

The available tasks in the SELANSI toolbox are listed and briefly summarized below.

- Definition of the problem:** The routine SELANSI_DataDef allows the user to easily define a new model (i.e. number of genes, interactions, type of kinetics and parameters), the simulation specifications (i.e. initial condition, time horizon for simulation and discretization including time and protein meshes), as well as the time steps for which the solution is saved.
- Modification of an existing problem:** SELANSI_Gnetmod and SELANSI_Meshmod allow the user to modify, respectively, the parameters and simulation specifications of an existing problem. Specifically, SELANSI_Gnetmod modifies the

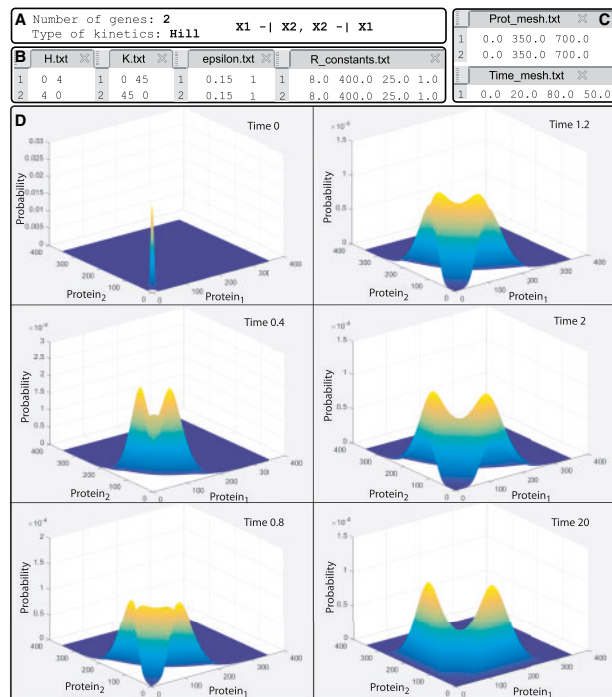


Fig. 1. Simulation with SELANSI of a two dimensional stochastic gene network with mutual repression and Hill kinetics. User inputs include (A) number of genes and kinetics, (B) kinetic constants, Hill coefficients and (C) protein and time meshes. (D) Samples of joint probability snapshots provided by SELANSI

default network parameters and feedback mechanism, while `SELANSI_Meshmod` modifies the default mesh for the semilagrangian method and the initial condition.

- iii. *Stochastic simulation:* `SELANSI_Solve` computes the (approximated) numerical solution of the CME obtaining the temporal evolution of the species' probability density function. It saves also the solution according to the user's specifications.
- iv. *Results and Visualization:* `SELANSI_Plot` depicts marginal and joint probability densities (in multidimensional case) for the time steps selected by the user.

As an illustrative example (this and other examples with different number of genes, connectivities and kinetics are provided with the toolbox) we consider a two dimensional gene regulatory network in which the two genes mutually repress each other. Figure 1 shows a scheme of the inputs provided by the user and the results of the simulation computed by SELANSI (for space reasons, we include only a few snapshots). For the selected values of the parameters the

network behaves as a transcriptional switch with a bimodal stationary distribution.

Funding

This work has been supported by Spanish MINECO grants AGL2015-67504-C3-2-R, PIE201230E0M2 (AAA), BES-2013-063112 (MP), MTM2016-76497-R, MTM2013-47800-C2-1-P (CV) and MINECO and European Regional Development Fund DPI2014-55276-C5-2-R (IOM).

Conflict of Interest: none declared.

References

- Dar,R.D. *et al.* (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. USA*, **109**, 17454–17459.
- Engblom,S. (2006) Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comp.*, **180**, 498–515.
- Friedman,N. *et al.* (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys Rev Lett.*, **97**, 168302.
- Gillespie,D.T. (2007) Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, **58**, 35–55.
- Hasenauer,J. *et al.* (2014) Method of conditional moments (MCM) for the chemical master equation. *J. Math. Biol.*, **69**, 687–735.
- Hegland,M. *et al.* (2010) CmePy documentation. <http://fcostin.github.io/cmepy/>
- Hespanha,J. *et al.* (2008) Moment closure for biochemical networks. In: *Proc. Int. Symp. on Communications, Control and Signal Processing*. pp. 42–147.
- Hoops,S. *et al.* (2006) COPASI – a COmplex PATHway Simulator. *Bioinformatics*, **22**, 3067–3074.
- Kazeroonian,A. *et al.* (2016) CERENA: ChEmical REaction Network Analyzer – a toolbox for the simulation and analysis of stochastic chemical kinetics. *PLoS ONE*, **11**, e0146732.
- Kryven,I. *et al.* (2015) Solution of the Chemical Master Equation by radial basis functions approximation with interface tracking. *BMC Syst. Biol.*, **9**, 67.
- Lapin,M. *et al.* (2011) SHAVE: Stochastic Hybrid Analysis of Markov Population Models. In: *Proc. 14th Int. Conf. on Hybrid Syst.: Computation and Control*. pp. 311–312.
- Maarleveld,T.R. *et al.* (2013) StochPy: a comprehensive, user-friendly tool for simulating stochastic biological processes. *PLoS ONE*, **8**, e79345.
- Pajaro,M. *et al.* (2017) Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting. *J. Theor. Biol.*, **421**, 51–70.
- Ramsey,S. *et al.* (2005) Dizzy: stochastic simulation of large-scale genetic regulatory networks. *J. Bioinform. Comput. Biol.*, **3**, 415–436.
- Sanft,K.R. *et al.* (2011) StochKit2: software for discrete stochastic simulation of biochemical systems with events. *Bioinformatics*, **27**, 2457–2458.
- Schnoerr,D. *et al.* (2015) Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.*, **143**, 185101.
- Thomas,P. *et al.* (2013) Intrinsic Noise Analyzer: a software package for the exploration of stochastic biochemical kinetics using the system size expansion. *PLoS ONE*, **7**, e38518.