Application of BioSPICE Model Definition Language To Stochastic Models

Chris D. Cox, J. Michael McCollum, Gregory D. Peterson
University of Tennessee

Timothy Elston, David Adalsteinsson
University of North Carolina

This document will describe the use of BioSPICE Model Definition Language (MDL), which at present is essentially Systems Biology Markup Language (SBML) level 2.0, to represent stochastic models of biological systems.

The standard approach for modeling biochemical networks is to derive ordinary differential equations (ODEs) based on the law of mass action for the concentrations of the biochemical species involved in the network. However, stochastic effects can be significant in cellular reactions, since the populations of molecules can be small. There are three basic types of stochastic models: 1) discrete models commonly solved by the Gillespie algorithm and 2) Langevin or stochastic differential equations (SDEs) in which the variables are approximated as continuous and a white noise term models stochastic behavior and 3) hybrid models which some chemical species are treated discretely while others are treated with the continuous approximation. The first method is exact but computationally slow, while the second and third methods provide excellent approximations under some circumstances and are computationally much faster. In all cases, the models are normally based on representing the biological system as a network of chemical reactions in which the reaction rate laws are governed by the law of mass action.

Applicability of SBML Level 2.0

SBML Level 2.0 fully supports Gillespie-type discrete stochastic models. Characteristics of discrete models include using the unit "molecules" for substances and the law of mass action for rate laws. Since the species are discrete, the variables are non-negative integers. Both of these conventions are fully supported by SBML level 2.0.

The chemical Langevin equations and hybrid models are derived from the Gillespie models. For hybrid models, it is necessary to indicate whether a chemical species is to be treated as continuous or discrete. We recommend that a new species attribute be added to support this declaration (e.g., discrete: boolean {use="optional" default="false"}). At present, this declaration can be handled as an annotation. The unit for both continuous and discrete substances is molecules; however, for the chemical Langevin equations, the substances are continuous rather than discrete random variables.

Relationship between ODE models and Stochastic Models

A number of model building tools are available both within and beyond BioSPICE for creating SBML. To date, many of the tools are designed to produce SBML for ODE models. In this section we will identify the key differences between deterministic ODE models and the various types of stochastic models to facilitate conversion between the two types of models, when appropriate. These differences are important to software integration in two ways. First, it is desired that model builder tools be capable of formulating both stochastic and deterministic models. Secondly, stochastic simulation engines should be able to interpret SBML for appropriate deterministic models in a stochastic framework. We will show below that the first task is much simpler than the second, and suggest MDL principles that will facilitate the latter.

First it is necessary to define which deterministic models are appropriate for interpretation in a stochastic context. First, it is those models that consist of networks of chemical reactions. While of diffusion processes have been represented stochastically (e.g., Stundzia, and Lumsden (1996) use the Gillespie algorithm to represent a diffusion-reaction process), at present the spatial aspects of SBML are not fully developed that would allow the work of determining the correct stochastic interpretation of a continuous diffusion model to begin. Secondly, mass action rate laws that are zero, first, or second order should describe the kinetics of the reactions in the network. Many ODE models combine several reactions of this type into a single reaction described by a more complicated rate law (e.g. Michaelis-Menten or Hill kinetics) using quasi-steady state (QSS) assumptions. While it is possible in some cases to make this assumption (Kepler and Elston, 2001; Rao and Arkin, 2003), in many cases stochastic simulations of reaction rate laws given by the QSS assumption will give incorrect answers. This topic will be discussed below, but for now the safest and best approach is to only consider ODE models composed of mass action kinetics.

Interpretation of an ODE model comprised of mass action rate laws for a Gillespie simulation is straightforward. For a Gillespie simulation, the initial substance amounts must be expressed in number of molecules. The easiest way to achieve this to specify these units at the time the model is defined and written in SBML. Therefore, to support stochastic simulations, model builder tools need to be able to specify substance amounts as molecules. The alternative is to take MDL written for a continuous ODE model and convert the initial amounts specified in units of concentration into units of molecules. This conversion requires the system volume to be known, so a minimum requirement for an ODE model to be interpreted stochastically is that the system volume must be specified. An added complication is that many different concentration unit definitions are possible in SBML. To create a model interpreter that could correctly handle all possible unit conversions would require a major effort. We recommend that BioSPICE MDL require a more limited set of units for ODE models that are meant to be interpreted in a stochastic context. For example, it could be required that all concentrations must be specified in moles/L and all system volumes must be specified in liters.

Secondly, the relationship between deterministic and stochastic rate constants must be considered (Gillespie, 1977). Consider the reactions:

$A + B \rightarrow C$ $k_{deterministic} = V \cdot k_{stochastic}$

$A + A \rightarrow A_2$ $k_{deterministic} = V \cdot k_{stochastic}/2$

Again, the best solution is for the stochastic rate constants to be entered by the user when using the model builder tool to write the SBML. The alternative solution is to have the stochastic simulation engine to automatically make the conversions. Again, a limited set of MDL units for models that are to be interpreted both stochastically and deterministically would greatly ease the task of writing the software to make this conversion.

One final consideration regarding Gillespie models is related to how rate laws are specified in SBML. For reactions governed by mass action kinetics the rate law is fully defined by the reaction definition and the rate constant. Therefore, there is no need to specify the rate law explicitly using MathML. The current practice of the University of Tennessee group when writing SBML is to specify the rate constant in the annotations that go with the reaction. We also include the complete MathML statement of the rate law to support the interpretation of our models by other groups. At present, we do not parse MathML. When reading our own SBML, we know where to look for the rate constant in the annotations. When reading SBML written by others, we define the kinetic law, but the rate constant would be undefined unless it has been included in the annotations using our format. Undefined rate constants would need to be specified manually prior to stochastic simulation. Ultimately one of two solutions needs to be implemented to promote software integration: 1) adopt a BioSPICE MDL standard of specifying the rate constant in the annotations for models that are to be interpreted stochastically or 2) require all SBML readers to parse MathML. The first solution is relatively simpler to implement than the second.

The chemical Langevin equations are a good approximation to Gillespie models when molecule numbers are sufficiently large. Surprisingly, this approximation can work quite well with molecules number of 20 or fewer. The chemical Langevin equations are stochastic differential equations (SDEs). When only elementary reactions are considered, the deterministic part of these equations is identical with their ODE counterparts. The strength of the noise terms in the SDEs is determined by the rates used in the Gillespie models. Defining chemical Langevin models requires the same information as the Gillespie models. That is, they require knowledge of the rate constants and stoichiometry of the underlying reactions. Therefore, the issues discussed above are also relevant to these models. The hybrid models, which combine discrete and continuous variables, require the user to specify the discrete and continuous groups of random variables. This could be added to the SBML specification without adversely affecting existing applications.

The software package Biochemical Network Stochastic Simulator (BioNetS) has been developed at UNC. BioNetS can simulate Gillespie models, as well as SDE and hybrid models. BioNets has a translator that maps SBML descriptions of biochemical reactions

to source code. The translator and source code generator are currently specific to Apples Computer's Mac OS X. However, efforts to make BioNetS platform independent are underway. The C/C++ code generated by BioNetS is portable and can be compiled and run in any computing environment.

Stochastic Simulation of QSS Rate Laws

As mentioned above, it is often possible to simplify deterministic models of biochemical networks using quasi-steady state approximations. The resulting model equations for the simplified system contain more complicated kinetics (e.g., Michaelis-Menten or Hill kinetics). Performing the same type of reduction on stochastic models is nontrivial and an active area research being pursued by both the University of Tennessee and University of North Carolina research groups. When these are developed, a more comprehensive ability to parse MathML may be required. At present QSS rate laws will generally not give correct results if used in a stochastic context; therefore, the present generation of stochastic simulation engines should indicate to the user that these types of models are not appropriate for stochastic simulation.

References

Gillespie, D.T. (1977), *J. Phys. Chem.* 81:2340-2361.
Kepler, T., Elston T. (2001). *Biophys. J.* 81:3116-3136
Stundzia, A.B. and Lumsden, C.J. (1996), *J. Comp. Phys.* 127:196-207.
Rao, C.V. Arkin A.P. (2003), *J. Chem. Phys.,* 118:4999-5010.