

SBML for discrete stochastic modelling, and the discrete stochastic models test suite

Darren Wilkinson

— Boston SBML Forum (October '05) —

School of Mathematics & Statistics,
University of Newcastle upon Tyne, UK
[also affiliated with Newcastle's *Systems Biology Resource Centre (SBRC)* and *Centre for Integrated Systems Biology of Ageing and Nutrition (CISBAN)*]

Discrete stochastic models in SBML

- Rate laws for discrete stochastic models represent reaction hazards, in units of # molecules / second (substance/time), and amounts are also measured in # molecules (substance)
 - Substance units — item

```
<unitDefinition id="substance">
  <listOfUnits>
    <unit kind="item"/>
  </listOfUnits>
</unitDefinition>
```
- Symbols in rate laws represent # molecules (substance, not concentration)
 - Species — hasOnlySubstanceUnits

```
<species id="X" compartment="Cell" initialAmount="0"
  hasOnlySubstanceUnits="true"/>
```
 - Not present in L1, so best to use L2...
- Not all model-building tools support non-default units or species with substance units...

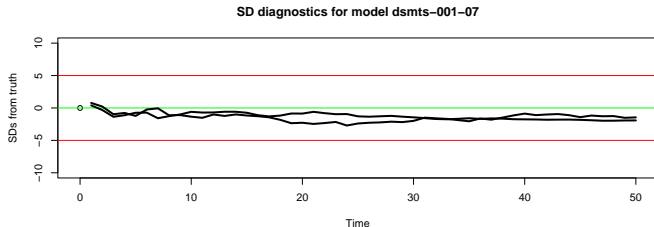
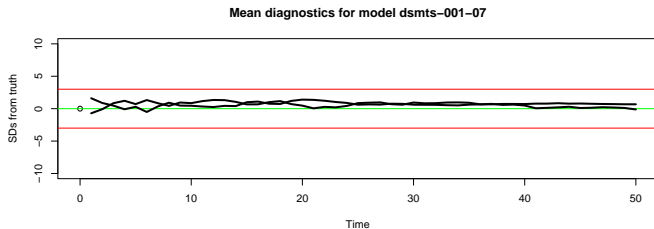
Testing stochastic simulators

- Stochastic simulators are harder to test than deterministic simulators, as they give different output on each run
- Idea: rather than testing the output at a collection of times, test the **distribution** of the output at a collection of times (obtained from a large number of simulator runs)
- Simplification: rather than testing for equality of the full distribution (not trivial), just test the first two moments (the mean and standard deviation)
- Use hand-crafted SBML models corresponding to models with tractable moments (birth-death, immigration-death, dimerisation kinetics)

Statistical tests for exact simulators

- Even if the simulator is perfect, the empirical mean and standard deviation will not exactly match the theoretical values
- Can use statistical tests to see if the extent of the deviation from the truth is “significant”
- Issues associated with multiple testing
- Exact test for mean. Approximate test for standard deviation
- Test chosen to ensure that false-positive rate is independent of the number of runs
- False-negative rate reduces as the number of runs increases (use 1,000 runs for a quick test, and 10,000 runs for a decent test)

Example test output



The test suite

Consists of:

- Web site including a list of FAQ and answers
- Overview document explaining the testing procedure
- Document describing the test models
- The test suite itself, consisting of:
 - Text file containing a list of model file names
 - Model files (in SBML L2v1 and SBML-shorthand)
 - Text (CSV) files containing correct mean and standard deviation at times $0, 1, \dots, 50$ for all model species
 - PDF plots of correct mean and standard deviation
- Some example code (in R) illustrating how to use the suite to test a particular discrete stochastic simulator (`gillespie2`)

What is covered?

- Current suite contains 22 models testing different aspects of simulator behaviour
 - Correctly encoded simulation algorithm, correctly discretised onto a regular time grid
 - Correct interpretation of volume, species symbols in MathML, `hasOnlySubstanceUnits`, amounts, rate laws, etc.
 - Correct interpretation of global and local parameters (including local parameters masking global parameters)
 - Correct behaviour with `boundaryCondition` (and `constant`) flag
 - Correct behaviour with a simple timed event and thresholded event
 - Includes models with zero, first, and second order mass-action rate laws

Other details

- Can also use the suite to test approximate stochastic simulators, but such simulators will fail the standard tests!
Can use less stringent tests such as: are all means and SDs within 3% of the true values, etc.
- The present release was put together by a Masters student (Tom Evans) under my supervision this summer
- Currently working on a new release (out before Xmas), that will include more models (testing more features, including non-mass-action rate laws, units, volume symbol in MathML, `initialConcentration`), and a major re-write of the documentation — but essential structure will remain identical

Other relevant work at Newcastle

- `gillespie2` — an SBML Level 2 discrete stochastic simulator that passes most of the tests
- `pysbml` — a python/libSBML-based tool-set for working with SBML models, including an interface to `gillespie2`
- `SBML-shorthand` — a concise notation for SBML documents, that is easy to read and write by hand, that includes support for discrete stochastic modelling features
- `BASIS` — Web/GRID portal including a repository of discrete stochastic SBML models relating to ageing

Current web home of the test suite:

<http://www.calibayes.ncl.ac.uk/Resources/dsmts/>



Kirkwood, T. B. L., R. J. Boys, C. S. Gillespie, C. J. Proctor, D. P. Shanley, and D. J. Wilkinson (2003). Towards an e-biology of ageing: integrating theory and data. *Nature Reviews Molecular Cell Biology* 4(3), 243–249.



Wilkinson, D. J. (2006). *Stochastic Modelling for Systems Biology*. Chapman&Hall/CRC Press. In press.

Contact details...

email: d.j.wilkinson@ncl.ac.uk

www: <http://www.staff.ncl.ac.uk/d.j.wilkinson/>