

## **Reviewer's report**

**Title:** The systems biology simulation core algorithm

**Version:** 1 **Date:** 10 March 2013

**Reviewer number:** 2

### **Reviewer's report:**

#### **SUMMARY OF PAPER**

The authors describe an algorithm for solving models encoded in SBML, as well as a Java-based implementation of the algorithm in the form of a software library, and finally, the results of tests on the library. The algorithm is based on a well-known and time-tested numerical integration algorithm by Rosenbrock, but can evidently support other integrators in a plug-in fashion.

As a reader familiar with SBML, but someone who has not implemented a simulator myself, I found the paper generally quite readable, pleasant and easy to follow, with helpful diagrams and flow charts.

#### **OVERALL ISSUES: MAJOR COMPULSORY REVISIONS**

1) The authors concentrate on differential equation models as if this modeling approach is the only one used in systems biology. This is certainly not the case: there is a variety of approaches used in systems biology, such as stochastic simulations, Boolean networks, rule-based modeling, etc. It is perfectly fine, and even desirable, to develop an efficient and complete ODE solver focused on SBML models (considering how popular SBML is and how complicated SBML models are to solve fully), but the authors should position their work in the proper context. I know for a fact that some of the coauthors are familiar with other simulation frameworks, so the problem is clearly one of how the paper is written and not a lack of understanding of the variety of approaches used in systems biology. Here are a few places where the text would benefit from revisions:

- The first sentence of the abstract: "With the increasing availability of high dimensional time course data for metabolites, genes, or fluxes the description of dynamical systems by differential equation models becomes more and more popular in systems biology." That is an assertion by the authors, and it reads as if it is a statement about the growth of popularity of ODE models compared to other techniques. Either the authors need to back it up with references showing that ODE modeling is gaining popularity compared to other methods, or else the statement should be modified.

- p. 2: "The models encoded in these formats can be interpreted in terms of differential equation systems, with additional structures such as discrete events and algebraic equations." They can be interpreted in other ways too.

- p. 3: "In order to derive an algorithm for the interpretation of SBML models ..."  
The statement should be qualified, for example "In order to derive an algorithm for the interpretation of SBML models in a differential equation framework ..."

There are probably other places where this issue should be addressed; the above are only some places where it was especially noticeable.

2) The section on Background on p. 2 makes statements about SBML and CellML, but the claims about CellML are overstatements. E.g., "... a large variety of programs supports these model description languages." Really? The authors want to claim there is "a large variety" of programs that support CellML? Another example: "Both communities offer standardized and manually derived benchmark test in order to evaluate the quality of simulation results." Benchmarks exist for CellML? If that's true, would the authors mind providing a reference for that claim? My impression has always been that CellML has nowhere the level of acceptance or testing that SBML has; therefore, to make such statements is not scientifically rigorous. Further, to any reader familiar with SBML, it comes across as a disingenuous attempt to make the authors' work seem more broadly applicable. The bulk of the text is intensely influenced by SBML, to the point where it uses terms such as "kinetic laws" and "algebraic rules" -- terms that do not exist outside of SBML, so the lip service paid to CellML support is not particularly believable anyway.

3) The description of how SBML constructs are handled is not complete. (Presumably this is a problem of exposition in the paper and not actually a failure of the implementation.) For example, SBML provides a facility for declaring new function definitions, but the authors do not describe how their algorithm would process such function definitions. The authors' software can probably handle them easily, but still, in a paper like this, it would be appropriate to include a few words of explanation about the stage at which function definitions are processed. Another matter is the ability in SBML to reassign stoichiometry values dynamically via the "StoichiometryMath" construct. Still another is whether the system does anything with SBML's "Constraints" construct. The description of the SBSCCL seems incomplete without some explanations, or at least passing mention, of these things.

4) There is very little discussion about the points of control that the user has to deal with. What are the settings on the numerical methods that have to be manually adjusted? (Or is the system so automatic that you can throw any model at it and it will figure out what to do?) A sentence or two would help readers to understand what they would be faced with if they tried to apply the library to a problem.

5) There is very little discussion about the limitations of the algorithm and software. It is implied that all possible SBML models can be handled, but are there any cases where (e.g.) the system has more difficulty converging on solutions than others? Are there other limitations?

## MINOR ESSENTIAL REVISIONS

p. 1:

- "or fluxes" -> "and fluxes".
- Comma after "fluxes".
- "becomes more and more" -> "is becoming more".
- In the last sentence of the first paragraph, "due to many special cases" is ambiguous. What has special cases? SBML or mathematical structures?
- "This article describes an efficient algorithm to interpret and solve differential equation systems in SBML models". This is not quite right. SBML models do not contain differential equation systems, or at least not unless they are structured in a very peculiar way that avoids the use of the SBML reaction and uses rules only. A statement such as "derived from SBML models" would be more correct.
- "pre-processing" does not need a hyphen.

p. 2:

- First paragraph: add "the" after "biology," and before "modeling".
- "efforts like this require a strong collaboration": remove "a".
- First sentence of 2nd paragraph: "or CellML" -> "and CellML".
- "online data bases" -> "online databases".
- "such as the BioModels database" -> "such as BioModels Database".
- "or the CellML model repository" -> "and the CellML model repository".
- "aproaches" -> "approaches".
- The sentence beginning with "For efficient model analysis" is very awkward and almost incomprehensible. Please rewrite.
- "benchmark test in order to" -> "benchmark tests in order to".

p. 3:

- "implicated" -> "implied"
- "further systems biological" -> "other systems biology" (note: 2 changes).
- "It comprises several" -> "It is comprised of several".
- "re-use" does not need a hyphen.
- "In order to derive an algorithm for the interpretation of SBML models": this is one of the places where the paper implies ODEs are the only option. This

sentence should be qualified, for example by being rewritten to say "In order to derive an algorithm for the interpretation of SBML models in terms of differential equations".

p. 4:

- "non-linear" does not need a hyphen.
- "Species, whose amounts": remove comma.
- "in any reaction and do hence have only" -> "in any reaction and hence only have".
- "A further concept of SBML models are" -> "A further concept in SBML is".

p. 5:

- "With the help of a bipartite matching" -> "With the help of bipartite matching".
- "Sub-systems" does not need a hyphen.
- "All kinetic laws of the reactions": the term "kinetic laws" is odd, and presumably something borrowed from SBML. Could the statement be qualified somehow to make that more clear?
- "Deviating initial values defined by initial assignments and are now assigned". The sentence doesn't make sense.

p. 6:

- "In order to deal with algebraic rules in an SBML model, these have to be converted to assignment rules." This is written as a universal statement, but is it not possible that a different approach might choose to do things differently? The authors should probably qualify that statement to say this is the approach taken in their particular algorithm.
- "these have to be" -> "they have to be".
- Omit the comma after "at least one variable,".
- "other equations in model" -> "other equations in the model".
- "the regarding algebraic rule": unfathomable.
- Omit the comma after "to computing a matching,".
- "An event in SBML is a function": no, it is not. It is a list of assignments.
- "In addition, a delay may": it is not clear what the source of the delay is. I think the authors mean that the SBML event concept has an optional slot for defining a delay; they are not talking about some external source of delays.
- "assignmens" -> "assignments".

- "exeuted" -> "executed".
- "but with an huge impact" -> "but these have a huge impact".
- Insert "For example," before "the order in which events".
- The statement "the order in which events have to be processed used to be at programmers discretion" is confusing. I think the authors mean that it was the case in SBML Level 2, and changed in SBML Level 3.
- "there are events, whose trigger has" -> "there are events whose triggers have" (3 changes).

p. 7:

- "non-persistent" does not need a hyphen.
- What is a non-persistent event? Is it an SBML concept?
- "This process repeats itself" -> "This process must be repeated".
- Omit the comma in "The precise calculation of the time, when events".
- Add a comma before "Rosenbrock's method determines the new".
- Add a comma after "that" in "After that the events".
- Omit the comma in "the time, at which".
- Change the colon to a period in "is calculated as follows:".
- Omit the comma after "in all dimensions, in which".
- Add a comma after "the second step" in "During the second step the derivatives of all species".

p. 8:

- There should be a reference given for the Apache Commons Math library.
- The 2nd paragraph talks about the use of KiSAO, but no explanation is given about the significance. What is it about? How is it used?
- "support for further community" -> "support for other community".
- "Biomodel 206" -> "BioModels Database model #206".
- "Biomodel 390" -> "BioModels Database model #390".

p. 10:

- "makes possible" -> "makes it possible".

p. 11:

- "a high performance of the overall library" -> "that the overall library offers high performance".
- Add comma after "SBML" in "models encoded in SBML returns".
- "In the SBMLInterpreter events" -> "In the SBMLInterpreter object, events"
- "currently stored in list" -> "currently stored in a list".
- Omit comma after "One could argue,".
- "where after" -> "whereafter".
- The last 2 sentences of that page are really unclear. I do not even know what to suggest.

Figures 2-4: I suspect the arrowheads on the arcs are too small for the final publication.

Captions: they all have more issues, but I did not have time to write them here.

#### DISCRETIONARY REVISIONS

1) There is very little discussion about the performance of their algorithm beyond some words about the speed with which it can solve test models and the models in the BioModels Database. The models in the SBML testing suite are apparently all relatively small. How does the system performance scale with model size? It would be useful to see some sort of plot or other comparison of (1) model size versus run time as well as (2) memory consumption versus model size. If available, a simultaneous plot against some other software they tested such as BioUML might be interesting (but not crucial).

2) SBML has a notion of extensions that add syntax to the basic definition of SBML. Some recent extensions include support for such things as decomposing models, and supporting rule-based models. It would be appropriate to mention (perhaps in the discussion section) how the library handles SBML models with these extensions, or what the authors' plans are for dealing with them in the future.

**Level of interest:** An article of outstanding merit and interest in its field

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.