Is your model VERIFIED?

Summary of interview with Dr. Herbert Sauro
By: Stacey Finley

What is the scope of the issue?
Approximately 6 years ago, researchers in the United Kingdom and New Zealand went about curating published systems biology models to determine how many of those can reproduce the results presented in the paper. Altogether, they found that more than 98% of the models were not reproducible based solely on the information published in the papers and supplementary materials. Instead, the researchers had to contact the authors of the original papers to get more information, or had to substantially alter the model to reproduce the published results. Unfortunately, this has become the norm when trying to implement a systems biology model published in the literature.

What are the reasons?
The natural question is why? Why are these models not reproducible? According to Sauro, the answers to this question are rather mundane, related to parameters, implementation details, or formatting.

- *Parameters*: the authors list the wrong parameter values, wrong units, or only a partial list of parameters.
- *Implementation*: authors provide an incomplete listing of the model equations or the wrong equations altogether, and fail to provide adequate details about how the model simulations were performed.
- *Formatting*: rather than providing the actual model file itself, model equations are provided in an appendix, and translating them into a simulation software introduces error.

Thus, there does not appear to be a conspiracy preventing those of us interested in a published model from reproducing the results given in the paper.

In addition, Sauro believes another reason is a lack of incentive for providing a reproducible model. Authors get no credit or acknowledgement if their model is actually reproducible. Thus, they have no motivation for doing so.

What is the impact of having reproducible models?
Saving another researcher the trouble of wading through model equations listed in an appendix is not a major outcome of having reproducible models. In the context of biomedical research, the significance of reproducibility in mathematical biology can have clinical impact. Sauro maintains that for large systems biology models, model construction must be more systematic. Testing should be treated as an engineering problem, where the assumptions and data used for model building and simulation are explicitly recorded. Only then can we produce clinically-relevant models that are trustworthy and reliable, such that they can be used to guide decisions in patient care.

What can be done?
As part of the CRBM, and in his own research group, Sauro is working to address this issue of lack of reproducibility in systems biology models. One role of the CRBM is to work with journals that are a natural home for systems biology modeling studies in order to increase the number of reproducible systems biology models that are published. This list of partner journals includes *Bulletin of Mathematical Biology*, *Biophysical Journal*, *Physiome*, and others.

The CRBM has established two ways in which it can work with journals to verify the simulated results provided in a manuscript.

1. *During peer review.* When a manuscript is submitted, the journal editor will list the CRBM as a reviewer, and the model will be checked to see if the results are reproducible. This option does not require any additional steps or new infrastructure for the journal editors or staff.
2. *After acceptance.* Models from papers that have been accepted will be checked prior to proceeding with the normal process required for publication.

If the modeling results can be reproduced, the model is categorized as having been verified by the CRBM. Additionally, if the model is provided in SBML or CellML format, the CRBM will annotate it to be more readable and user-friendly (i.e., replacing all of those generic "*x*" variables with more descriptive names!). The Center is starting a pilot program soon, in which this model verification service will be free for five years.

What is the long-term solution?

This issue of reproducibility needs broader and more long-term support, both from funding agencies and from the scientific community.

Sauro believes that in general, the research enterprise in the United States has not invested in the resources needed to support reproducibility in mathematical biology modeling. In contrast, European institutions have both the infrastructure and the mandate to promote reproducibility.

However, even without funding from federal sources, the hope is that over time, the importance of reproducibility in mathematical biology modeling will permeate through the research community. "This is just something that is needed," Sauro states. He likens it to addressing grammar and readability in a manuscript submitted for peer review. Just as authors spend time and effort to remove the grammatical errors in the writing, they should check that the model provided with the manuscript is error-free and generates the results in the paper.

The burden is also on journal reviewers and editors, who can promote a paradigm shift. Reviewers often mention, either directly to the authors or in comments to the editor, if grammatical errors are substantial enough to be distracting and make the paper unreadable. So too should they comment on whether it is possible to reproduce the manuscript's results using the model provided by the authors. Journals can make this part of the editorial decision. In addition, a badge of verification listed on the journal's website can serve as a way to encourage authors to make reproducibility a priority.

Going even further, the CRBM is also interested in using PubMed as a platform to acknowledge if a published model is reproducible, an ultimate verified status. Interestingly, papers archived in the Europe PubMed Central link to BioModels, a curated repository of mathematical models in biology and biomedical research. This is another case where the U.S. is trailing behind other entities in supporting reproducible science.

Concluding thoughts

The lack of reproducibility in mathematical biology can have far-reaching effects. As a scientific community, we must be committed to reproducibility in our modeling. Furthermore, journal editors and reviewers should hold researchers accountable for providing reproducible models. The CRBM is poised to help make substantial strides in this regard and in improving best practices in modeling.