Editor's review: "automan: a Python-based automation framework for numerical computing"

Lorena A. Barba[1]

[1]George Washington University

April 28, 2020

Summary

This is the editor's review of manuscript CiSESI-2018-02-0013, submitted to *Computing in Science & Engineering*, Reproducible Research Track: *"automan: a Python-based, automation framework for numerical computing"* (Ramachandran, 2018).

We received two peer-review reports for this manuscript, and the reviewers were: Thomas Pasquier and Olivier Mesnard. Both reviewers recommended that the manuscript be accepted with a minor revision, and I agree.

Reviewer 1 comments on software-engineering aspects: he asks if `automan` could be combined with continuous integration tool chains. He also asks to state plainly the license of `automan`—from the LICENSE.txt file in the GitHub repository, it looks like it's under BSD3, but you could state this in the README and in the paper.

Reviewer 2 acted as Reproducibility Reviewer. He inspected the code repository of `automan`, installed it, and tested it with a fabricated problem. He also provides several minor suggestions in his open report (Mesnard, 2018).

Editor's review

One issue that the author needs to correct is the terminology. The CiSE RR Track adopted a set terminology that we request authors to consistently apply. See the Call for Papers (Barba and Thiruvathukal, 2018a).

Citing from there:

> Reproducible research is defined as research in which authors provide all the necessary data and the computer codes to run the analysis again, re-creating the results. Replication, on the other hand, is defined as arriving at the same scientific findings as another study, collecting new data (possibly with different methods), and completing new analyses.

Throughout the manuscript, the author is using "repeatability" and "repeating" in place of "reproducibility" and "reproducing." Conflicting terminology is a source of confusion in the field, and the CiSE RR editors want consistent usage in the magazine.

1

I also recommend reading: *"Re-run, Repeat, Reproduce, Reuse, Replicate. . . "* (Benureau and Rougier, 2018)

Abstract

CiSE articles don't normally carry a standard abstract. Instead, a short "blurb" appears under the title, usually about three sentences long, to pull focus. Please edit the abstract down, in line with this style. Here's a starting point for your editing (or simply re-use this, if you like it):

> Automating computational workflows leads to more reproducible research. The Python tool `automan` assembles a suite of long-running computations, deploys them to your computational resources, and produces final plots from output data, all with a single command.

Introduction and Discussion

In the Introduction, you say "What is perhaps more important is that we discovered that [reproducibility] can be very profitable to the researcher." The paper then goes into the design details and explains the usage of `automan` in section 3, and you wrap up with a summary in the first paragraph of section 4 (Discussion). Towards the end, the Discussion explains how `automan` helped you manage a series of about 75 computational experiments with SPH schemes. But I miss here a stronger, more persuasive statement about how `automan` expedited and ensured quality of the research (it doesn't help that the paragraph in question is plagued with passive phrases). I would like it if you re-wrote this passage to reveal more about your own experience, and how you benefitted from the tool you created. Convince your readers here that the initial overhead of setting up `automan` for their simulation-based workflow will pay off!

Editorial / style requests

The manuscript is well written, overall, but a few small improvements would polish it further. In CiSE, we aim for a more readable style than typical scholarly literature. In line with that, I want to encourage you to especially look at whether any passive constructions are strictly necessary, and look to remove syntactic-expletive phrases like "there are," if you can, as well as superfluous adverbs and superlatives (e.g., clearly, very).

Examples:

First sentence! "It is well known that. . . " You can easily remove that, and combine the first two sentences: "Reproducibility is a cornerstone in all science fields, and is vitally important in computational science."

p. 1, l. 39: "Unfortunately, there are not many immediate or direct incentives for a researcher to invest time. . . " >> "Unfortunately, researchers have few immediate or direct incentives to invest time. . . "

p. 1, l. 41: "there can be significant challenges involved in [reproducing] and replicating numerical experiments" >> "reproducing and replicating numerical experiments involves serious challenges" (or another adjective, but don't use "significant" unless you mean statistically significant, to avoid misunderstanding).

p. 1, l. 45: "we believe this is a very important step in facilitating reproducible research" >> This step is important for facilitating reproducible research" (the terminology here is consistent with CiSE usage).

p. 1, l. 46: "The approach used in our automation framework is fairly general" >> "Our automation framework typifies a general approach"

p. 1, .. 49: "[reproducibility] can be very profitable to the researcher" (remove "very")

p. 1, l. 51: "It is our hope that" >> "We hope that"

— The fact that we can pick several examples from just the first two paragraphs shows that you could polish the manuscript by combing through the text looking for just these patterns. I expect about a 5–10% reduction in word count from this exercise!

p. 10, l. 40: Bad sentence: "It is important to note again that the commands that are executed such that one may configure the directory in which they will generate output using a command line argument."

Typos:

Title: you don't need the comma.

p. 1, l. 38: "some of [the] most important articles" (missing "the")

p. 3, l. 21: capitalize Docker

Software Appendix

As indicated in the Call for Papers: "articles will include an appendix reporting on the software engineering and data management practices followed by the authors." Please add a description to that effect, in the form of a brief Appendix.

Extended Acknowledgement

In the short statement on *"Peer review in the CiSE RR Track"* (Barba and Thiruvathukal, 2018b), we describe our ambitions for opening up the peer-review process in this track. Your reviewers opted in to open identity and open reports. We ask that you recognize reviewers by name in an extended acknowledgements section, briefly stating their contribution to your paper. You can cite the open reports with the URL or DOI, in each case.

References

Please provide DOIs wherever possible for all references.

For Ref. [1], please add DOI of the PDF version of this blog post, published on Figshare (10.6084/m9.figshare.4879928.v1).

For Ref. [8], please update the citation, as needed.

References

L. A. Barba and G. Thiruvathukal. *Reproducible Research Track – Call for Papers - Computing in Science & Engineering — IEEE Computer Society.* https://publications.computer.org/cise/2017/07/26/reproducible-research-track-call-for-papers/, 2018a. URL https://publications.computer.org/cise/2017/07/26/reproducible-research-track-call-for-papers/. Accessed on Sat, June 09, 2018.

Lorena A Barba and George K Thiruvathukal. *Peer review in the CiSE RR Track*, 2018b. URL https://doi.org/10.22541%2Fau.152553544.46681118.

Fabien C. Y. Benureau and Nicolas P. Rougier. *Re-run Repeat, Reproduce, Reuse, Replicate: Transforming Code into Scientific Contributions*. Frontiers in Neuroinformatics, 11, jan 2018. doi: 10.3389/fninf.2017. 00069. URL https://doi.org/10.3389%2Ffninf.2017.00069.

O. Mesnard. *Review: automan: a Python-based, automation framework for numerical computing*. https://www.authorea.com/users/99991/articles/307464-review-automan-a-python-based-automation-framework-for-numerical-computing, 2018. URL https://www.authorea.com/users/99991/articles/307464-review-automan-a-python-based-automation-framework-for-numerical-computing. Accessed on Sat, June 09, 2018.

P. Ramachandran. *[arxiv:1712.04786] automan: a simple, Python-based, automation framework for numerical computing*. https://arxiv.org/abs/1712.04786, 2018. URL https://arxiv.org/abs/1712.04786. Accessed on Sat, June 09, 2018.