

Review: Reproducing the Results for NICER Observation of PSR J0030+0451 by Afle et al.

By Line Pouchard, Brookhaven National Laboratory
Computing in Science and Engineering, Reproducibility Track
DOI:10.5281/zenodo.8034613

Major comments:

The paper “Reproducing the Results for NICER Observation of PSR J0030+0451” by Afle et al. and submitted to the CISE Reproducible Research Track describes an experiment where a team who is not part of the original analysis reproduces results derived from observations of a pulsar. The authors calculate the mass and equatorial radius of this pulsar with a computationally intensive equation of state and succeed in obtaining statistically consistent results in the reproducibility experiment. This experiment shows that the same software stack as in the original paper can be re-used and expanded: this is an innovative contribution for reproducibility in computational science. Attempts to reproduce similar astrophysics analysis without using the same software as the original team showed that results were replicable but not reproducible (a different computational software was used in previous literature to produce comparable results). Team members from their previous reproducibility paper had participated in the original and reproducibility experiments.

Afle et al. provide a thorough description of their challenges and lessons learned in reproducing this analysis, and give an annotated inventory of the data, software code, documentation, and configuration files available with the original paper. They note that missing post-processing and plotting scripts prevented them from reproducing figures, but they were successful in reproducing computational values.

The re-use of the computational software stack enables another innovative contribution in this paper. Using Bayes theorem to infer the posterior probability distribution for the same parameters as the original analysis, Afle et al. demonstrate that they obtain the exact measurements as the original analysis within the 68th percentile. In addition, the authors’ original analysis is expanded by broadening the Bayesian priors for the pulsar radius from 16 to 25 km. This expansion shows that the reproducible results are robust to the choice of priors in a Bayesian paradigm.

Afle et al provide a docker container that produces an image of the code that the authors used to verify their reproducibility analysis. The docker container provided in footnotes can be downloaded and installed. While the original work used the Dutch national supercomputer Cartesius, the authors show that the reproducibility analysis can be performed on an another commodity cluster, here the Syracuse University Gravitational Computing Cluster. This is also an important outcome for reproducibility research where accessibility to the appropriate systems is required.

This paper and the experiment it describes contains lessons learned for reproducibility research and points to expanding research directions. No matter how well intentioned the original authors are in sharing their data and code, some pieces may end up missing. Providing a Docker container for the re-analysis is useful but not sufficient as the authors needed to derive and customize Singularity images from their Docker container to deploy on a managed system.

Case studies on reproducibility such as this one are useful to explore the breadth and depth of this research topic in the physical sciences and I recommend acceptance of this manuscript. While very valuable as a case study, it is still up to the reader to extract a generalizable method from this paper for reproducibility. One can start by pointing to the requirements demonstrated here:

- a) a non-trivial amount of work is required to reproduce such results, and publishing them is worthy of such efforts.
- b) details of what an original author needs to make publicly available to ensure reproducibility, including a full computational stack, are indispensable
- c) access to adequate computational resources is needed, and
- d) statistical analysis showing consistency between the original and reproduced outcomes improves the contributions of the reproducibility paper.

The paper shows that, in facilitating reproducibility, the lack of formal methods or at least a standard code- and data-sharing framework and best practices employed by original authors may result in missing pieces and impede attempts at reproducibility. For an approach towards the development of uncertainty-aware metrics of reproducibility, see <https://arxiv.org/abs/2301.05763>.

Minor Comments:

1) The authors may want to cite the National Academies of Sciences, Engineering and Medicine (NASEM) report for definitions of reproducibility and replicability:

Reproducibility and Replicability in Science. 2019. doi: [10.17226/25303](https://doi.org/10.17226/25303).

The Computing in Science and Engineering journal uses a definition congruent with the NASEM report: "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."

2) There are words missing on P4 line 53 2nd column, and P9 line 35 -36 2nd column

3) Suggesting additional ACM keyword is requested by the CISE review form, this reviewer would recommend adding these keywords from ACM taxonomy:

D2.1.e: Software engineering methodologies

D.3.3.h Distributed objects, components, containers

Acknowledgments: This review has been created by employees of Brookhaven Science Associates, LLC operator of Brookhaven National Laboratory, a U.S Department of Energy Office of Science laboratory operated under Contract No. DESC0012704