Assessing Quality of Code in Replication of Physics Research

Alexander Dickinson, Matthew Krafczyk, and Victoria Stodden
National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

Introduction
The replication of results is essential to the scientific method, as it helps prevent fraud and data manipulation. This also instills confidence in science. Though the information for replicating research can generally be found in the methods and materials section of a paper, computational research requires what was described by Stanford researcher Jon Claerbout as “really reproducible research,” which includes the “complete software development environment which generates the figures.” So computational replicability is seriously hampered by missing or poorly-written code.

Research into computational replicability has focused on availability of code thus far. The results are not encouraging. A question that remains is whether provided code can actually be used to replicate results.

Aim
I will attempt to compile and run code used in physics research, compare any results to those published in papers, and document the entire process, to answer the following questions:

1. Will code be platform-dependent?
2. How will the quality of documentation affect usability?
3. Given compilation or runtime errors, how difficult will code be to debug? How feasible is this?
4. Is there an expectation that users have the appropriate knowledge of physics to reproduce paper results?

Method
Attempts will be made to replicate the results of as many papers as time permits. The process of compiling and running code will be documented. If code is compiled successfully, output will be compared to figures in the corresponding paper. If results bear some resemblance to figures in the paper, but domain knowledge is required to reproduce results, I will move on to the next paper.

Methods for compiling and running code will vary depending on the code provided.

For each paper, notes will be compiled for the following stages of replication:
1. Building source code.
2. Running built code.
3. Comparing produced output to article results.

Results
Though the quality of provided code and documentation has varied, researchers have so far failed entirely to provide code that replicated results in a paper. Some problems could potentially be resolved given communication with the authors and more domain knowledge.

Of the eight papers I have looked at, three have provided code that might replicate the paper’s results. Code for these papers had no accompanying documentation, and some of the code had to be rewritten for compilation to succeed. The code for two of these papers produced output that clearly had some resemblance to the paper’s results. The code for the last of these three papers is still being debugged.

The rest of the papers provided libraries. Each library had code that was related to testing. Tests were not relevant to the specific results of corresponding papers.

Example Output
This is a grid used to demonstrate a method for solving a certain differential equation. This represents one of the only diagrams that have been produced by code. It is not quite accurate, but more work might yield the correct diagram.

Conclusions
Code used in computational physics papers is not written with replication in mind. Researchers should not see code as just a tool but as an important part of their research. Code and scripts that make replication simple should be the norm.

Acknowledgments
I would like to thank Matthew Krafczyk and Victoria Stodden for their assistance in my research. I would also like to acknowledge the National Science Foundation's funding of REU INCLUSION.