

Panel discussion: Reproducible research

Panelists: Hannah Christensen, Julia Keller, Linus Magnusson, Jenny Sun

Chairs: Georgios Fragkoulidis, Sebastian Lerch, Stephan Rasp

Introduction to computational reproducibility

Sebastian Lerch

Conference on Predictability and Multi-Scale Prediction of High Impact Weather, Landshut, October 2017

Heidelberg Institute for
Theoretical Studies



Disclaimer

These slides are partly based on presentations¹ by Victoria Stodden, one of the leading scientists working on open data, reproducible research, and tools for implementing reproducibility.

¹<http://web.stanford.edu/~vcs/Talks.html>, particularly <http://web.stanford.edu/~vcs/talks/FRA2017-STODDEN.pdf>

Reproducibility crisis and the credibility of science

[—]

Shifting paradigms in science, and reproducibility concerns

- ▶ Over the past decades, computational methods have radically changed science.
- ▶ Virtually all discoveries have a computational component (processing and analyzing data; simulation of complex systems, ...).

Shifting paradigms in science, and reproducibility concerns

- ▶ Over the past decades, computational methods have radically changed science.
- ▶ Virtually all discoveries have a computational component (processing and analyzing data; simulation of complex systems, ...).
- ▶ Mismatch between traditional scientific process and computation leads to reproducibility concerns:
 - ▶ Lack of transparency and disclosure in computational methods makes it almost impossible to reproduce and verify published results.
- ▶ Facets of reproducibility
 - ▶ reproducibility as independent third party
 - ▶ reproducibility as researcher conducting the work

Reproducibility in computational science

David Donoho, 1998, inspired by Jon Claerbout:

The idea is: An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures.

Two types of reproducibility, both are required:

- ▶ reproducing the computational steps
- ▶ re-implementing the experiment independently

Importance of reproducibility

- ▶ One of the main principles of the scientific method, and elementary to scientific advancement.
- ▶ 'The success and credibility of science are anchored in the willingness of scientists to ... expose their ideas and results to independent testing and replication by others. This requires the open exchange of data, procedures and materials.'
(American Physical Society)²
- ▶ Problematic implications of erroneous results in evidence-based decision making: policy, medicine, ...
- ▶ Journal policies and funding agency mandates.

²https://www.aps.org/policy/statements/99_6.cfm

Researchers' concerns and practical issues

- ▶ investment of time and effort that could be spent otherwise
 - ▶ code is usually not well designed or documented
- ▶ current scientific culture does not reward producing reproducible research
 - ▶ lack of proper attribution
 - ▶ potential loss of future publications
 - ▶ potential disadvantage to competitors
- ▶ legal and practical issues in sharing data and code
 - ▶ data policies (e.g. those of meteorological services) may not allow sharing data
 - ▶ how to publicly store a possibly TB-sized data set?
 - ▶ code may depend on proprietary software