PERSPECTIVE • OPEN ACCESS

The critical need to foster computational reproducibility

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

PERSPECTIVE

The critical need to foster computational reproducibility

Robert Reinecke[1,2,*] , Tim Trautmann[3] , Thorsten Wagener[1] and Katja Schüler[4]

1 Institute of Environmental Science and Geography, University Potsdam, 14476 Potsdam, Germany
2 International Centre for Water Resources and Global Change (UNESCO), 56002 Koblenz, Germany
3 Institute of Physical Geography, Goethe University Frankfurt, 60438 Frankfurt, Germany
4 University Hospital Frankfurt, Goethe University Frankfurt, 60438 Frankfurt, Germany
* Author to whom any correspondence should be addressed.

E-mail: robert.reinecke@uni-potsdam.de

1. Introduction

The climate crisis illustrates the need for earth and environmental models to critically assess the planet's past and future by translating emissions into climate signals and subsequent impacts regarding floods, droughts, heatwaves, and other hazards. While computational models grow in relevance by guiding policies and public discourse (Saltelli *et al* 2020), our trust in these models is put to the test. A recent study estimates that 93% of hydrology and water resources publications cannot be reproduced (Stagge *et al* 2019). In this perspective, we question whether we are amid a reproducibility crisis in the computational earth and environmental sciences (shortened to earth sciences in the following) and take a peek behind the curtain of everyday research.

Software development has become an integral part of research in many areas (Virtanen *et al* 2020), including the earth sciences, where computational models and data processing algorithms become increasingly sophisticated to solve the challenges of our time, like simulating the impacts of a changing climate. Paradoxically, this development threatens scientific progress: reproducibility, an essential pillar of science, is increasingly difficult to reach as software and data are often inaccessible (Añel *et al* 2021, Hutton *et al* 2016, Peng 2011). While retracing results through independent implementations is important, access to the original computational experiment is key to understanding critical explicit or implicit assumptions and their effect on the experiment's results (Stodden *et al* 2016).

This trend is particularly worrisome as scientific results have potentially controversial implications for stakeholders and policymakers and may influence public opinion and decisions for a long time (Munafò *et al* 2017). In recent years, progress towards Open

Science (Hall *et al* 2022) and the implementation of FAIR (findable, accessible, interoperable, reusable) principles (Wilkinson *et al* 2016) has led to more publishers demanding access to data and source code alongside peer-reviewed manuscripts (e.g. GMD 2019).

We argue that we insufficiently understand how the earth science community attempts to reproduce computational results and what challenges they face in this effort. To what do scientists attribute this lack of computational reproducibility, and what are possible solutions?

To lay a path for a future where Open Science is the norm, we let the community speak on what they think is necessary and paint a picture of a future that fosters computational reproducibility and thus trust. Our non-representative poll through a web-based survey revealed that: (a) the lack of reproducibility is jeopardizing trust in computational research, (b) that there is a considerable lack of knowledge on establishing software development methods, and (c) that Open Science is still not widely practiced.

2. Methods

Following established standards on the design of polls, questions were composed based on a list of initial hypotheses (supplement available online at stacks.iop.org/ERL/17/041005/mmedia) in multiple brainstorming sessions. The initial set of questions was then integrated into a polling tool (soscisurvey.de) and pretested with a selected focus group of ten researchers not involved in the study's design, including PhD students, three professors, postdoctoral researchers, and one head of a research institute. Half of the focus group consisted of researchers with a global modeling background; others had varying

regional backgrounds in hydrology. Their feedback was used to revise the initial set of questions.

The poll features 21 questions grouped into four categories: demographic information, opinion, behavior, and solutions (supplement). The poll is biased towards scientists from the hydrology (figure S2) community due to the scientific background of the authors and their access to distribution channels (see supplement for biases and sampling limitations).

Our definition of reproducibility was stated prominently at the beginning of the poll. Further, it was added as a 'tooltip' to every mention of the term to ensure that participants would always relate their answers to our definition. Definitions of the term reproducibility differ broadly among scientists and fields (e.g. Goodman *et al* 2016; www.acm.org/publications/policies/artifact-review-and-badging-current). Here we follow Stodden *et al* (2016) in their assessment that without being able to redo the exact experiment (*methods reproducibility* Goodman *et al* (2016)) we cannot advance our science because we will not be able to assess which assumptions may have led to the outcomes (*results reproducibility* Goodman *et al* (2016)). Thus, for the presented poll, we chose to define reproducibility as: 'results obtained by a modeling experiment should be achieved again with a high degree of agreement when the study is replicated with the same model design, inputs, and general methodology by different researchers. We explicitly exclude the retracing of results by means of using a different modeling environment (including variations in model concept, algorithms, input data or methodology)'.

A total of 347 participants from multiple fields within the earth sciences participated (figure S2) during 2 months in spring 2021. A large number of 265 participants completed all of the poll questions and were included in the shown analysis. All plots can be automatically reproduced from the raw data in the supplemental material.

3. The lack of reproducibility is jeopardizing trust in our results

The poll shows that reproducibility is a topic in the computational earth sciences that urgently requires our attention. Only 3% of participants strongly agree that most science in their field is reproducible, while the majority disagrees, highlighting that reproducibility is a significant issue which jeopardizes the trust in our computational science (figure 1). This perception shows no correlation to the career stage. However, the results also show a statistically significant discrepancy between how researchers see themselves and how they perceive others in their field: the majority believes their own science is reproducible and agree that they themselves could teach reproducibility methods (figure 1).

Overall, 59% of all participants never ran somebody else's model to reproduce their results, and 48% never used their own model to attempt reproducing the conclusions derived from other models. These results align with previous polls' findings that did not focus on the earth sciences or specifically any computational sciences (Baker 2016). However, while the broader scientific community regarded selective reporting and pressure to publish as the main reasons for lack of reproducibility (Baker 2016) (reasons also mentioned by our participants), our results highlight that reasons are manifold and possibly specific to computational sciences.
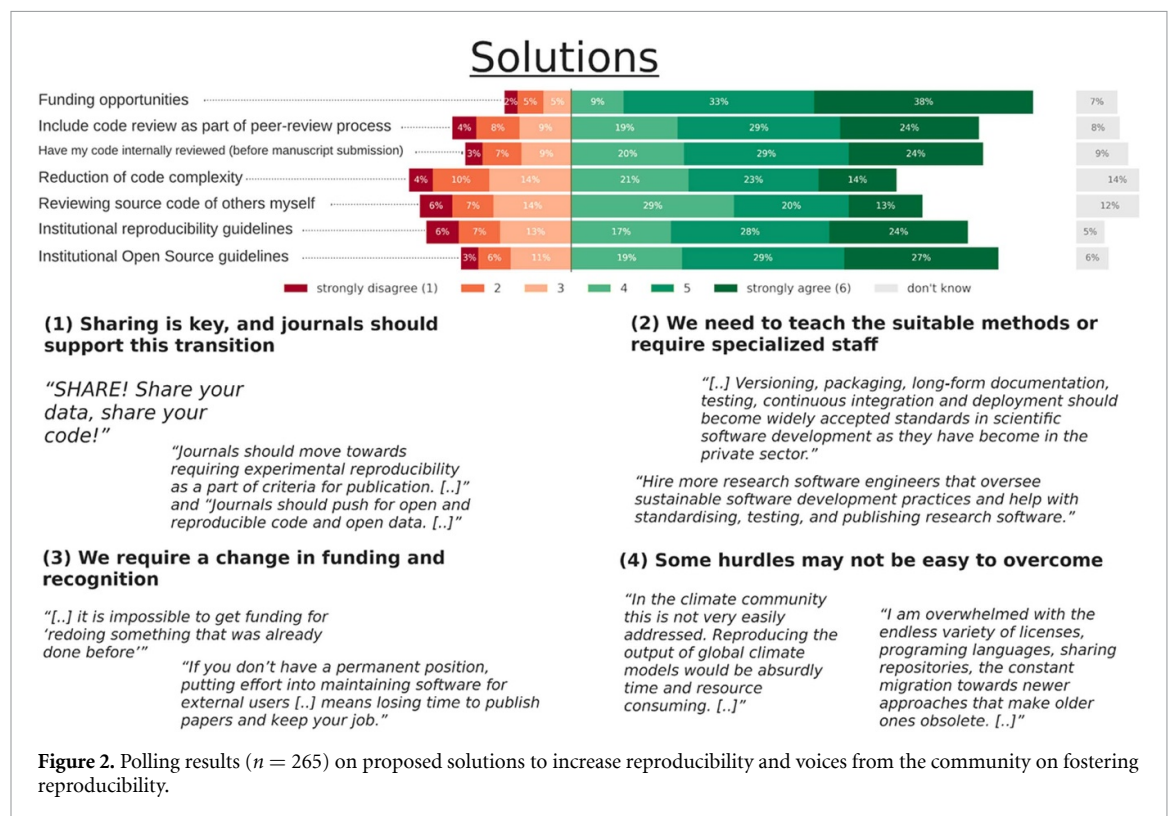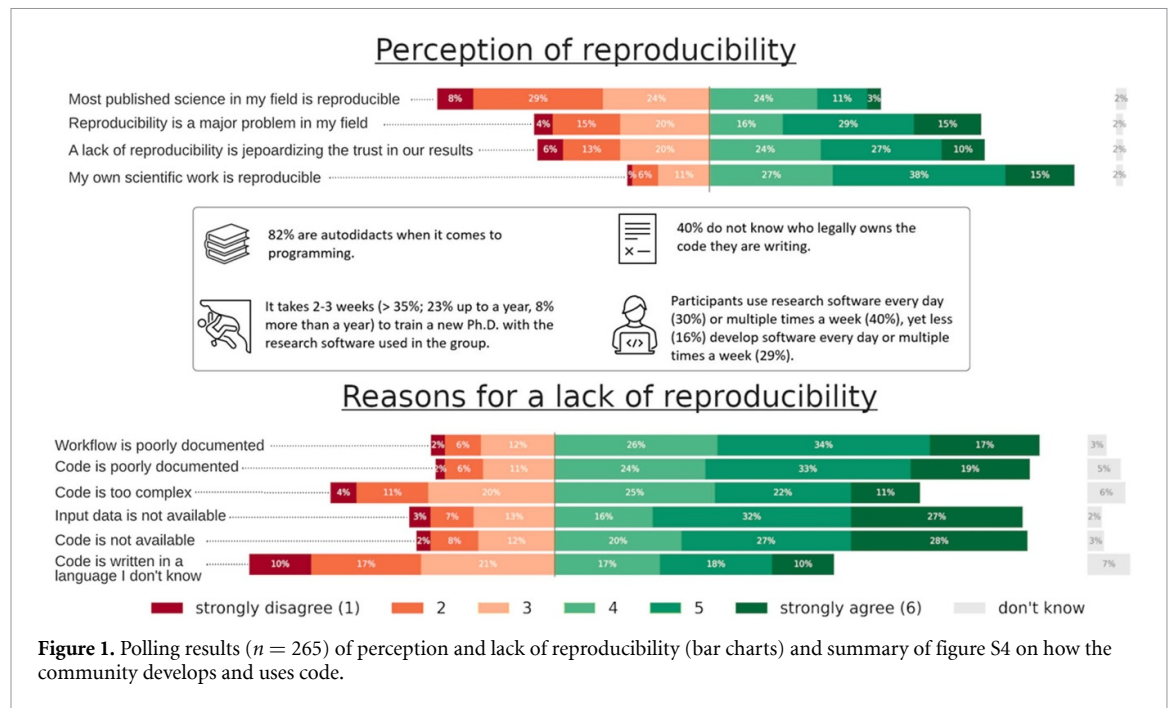
A majority of our poll participants agreed that they lack resources to improve reproducibility more than the necessary knowledge (figure 1). From the available options, a poorly documented workflow (77%) and a lack of code documentation (76%) showed the highest agreement among participants as a reason for the lack of reproducibility, followed closely (75%) by code and data availability.

If code and input data availability is one of the main reasons for the lack of reproducibility in the earth sciences, what keeps researchers from publishing their code? Twenty-nine percent of participants already publish all their code as Open Source, while 33% said that the main reason for not publishing their code is a lack of funding. Only 9% mentioned the fear of losing the lead on other groups not to publish their code. Other reasons were a fear that the code is too complex to understand for others, poorly documented, a general supervisor opposition, or shyness and doubt if the code is helpful to others.

4. Perceptions of the research community: paths towards reproducibility

What are the paths to increase reproducibility in the computational sciences? According to our poll, a solution can be found in one of the causes. Increased and specific funding is mentioned as a leading answer to the current lack of reproducibility (figure 2). This differs from an extensive poll among scientists from various fields which did not focus on computational earth sciences where 'better understanding of statistics' was a primary factor that could boost reproducibility (Baker 2016). Clearly, solutions to increase computational reproducibility differ from solutions for empirical research from other fields.

While funding received the highest agreement (80%), other proposed solutions such as institutional Open Source guidelines (75%), internal review (73%), or including code review in the peer-review process (72%) were also positively rated (figure 2). Additionally, scientists participating in the poll provided their own solutions in an open text field. The proposed solutions can be clustered into four major categories discussed in the following with

Figure 1. Polling results ($n = 265$) of perception and lack of reproducibility (bar charts) and summary of figure S4 on how the community develops and uses code.

Figure 2. Polling results ($n = 265$) on proposed solutions to increase reproducibility and voices from the community on fostering reproducibility.

exemplarity quotes from the poll shown in figure 2 (detailed information in the supplement).

A significant hurdle for reproducibility is a lack of code and input data (figure 1). Thus, sharing code and data is one of the most mentioned solution to increase reproducibility. Scientific journals are a significant puzzle piece in supporting this transition (figure 2).

Knowledge about software development methods is limited as most earth scientists are self-taught programmers (figure 1). As a result, multiple poll participants call for increased efforts in teaching software engineering methods (figure 2). On the other hand, participants also argue that earth scientists are not software engineers, and thus the solution is to hire specialized research software engineers. This is also discussed in Hut *et al* (2017).

Another aspect of increasing reproducibility is funding opportunities, as the continuous call for innovation leaves no room to reinforce or question existing knowledge (figure 2). A parallel issue

is the pressure to publish in the academic career system.

While most answers can be found in the first three categories, other researchers see obstacles that are not easily overcome, for example, extensive computational requirements in specific fields like climate modeling (figure 2). Others see a challenge in progress itself and feel overwhelmed.

5. Perspective

Reproducibility and open science are gaining increasing attention (e.g. Hall *et al* 2022), and all over the world, initiatives for open science and reproducibility like the Reproducibility Network (ukrn.org/international-networks) and dedicated journals (rescience.github.io, joss.theoj.org) are founded. However, reproducibility is challenging and requires additional resources to document processing steps and assumptions (Thornton *et al* 2005). Even reproducing one's research from the past can be very difficult (Perkel 2020). While the poll participants agree that computational research is currently not reproducible, the majority of researchers are keen on improving beyond the current state.

The results presented here are an explorative snapshot on this topic and possibly not representative for the whole earth science community. Nevertheless, we deem the results crucial and urgent enough to be discussed in a broader context.

To progress, wider availability of code and input data is crucial. However, this requires resources to be supported through existing or novel funding frameworks (Knowles *et al* 2021). Such changes require systematic alterations to how we attribute academic success. Time spent on these issues will limit writing papers. We require adequate acknowledgment of the additional work required to publish reproducible science and thus good research software.

Most researchers are autodidacts when developing software and lack the necessary knowledge on industrial standard code development methods and software licensing. Well-tested industry-standard code is assumed to have 10–15 bugs per 1000 lines of code (McConnell 2016). We can only speculate what this frequency looks like for research code, with dire consequences for the validity of our research. Access to software engineers through universities and institutions is a possible counter-measure that should be encouraged and could ultimately lead to a more robust software and better and more impactful research (Hut *et al* 2017).

Multiple solutions have been proposed to tackle the outlined issues. Using software containers (technology that bundles an application together with all requirements, such as libraries, for running the software) to distribute models to different computing environments allows for easier access to experiments and dynamically compiled documents

(Nüst *et al* 2020; supplement for additional references); automated executions of the code during peer review process allow for a more transparent connection of experiment and publication (codecheck.org.uk/project). Some journals have taken the step of enforcing rules on code and data availability. Others have started to find synergies in teaching postgraduates from different fields in joint programs or written guidelines for their respective communities for Open Science.

As the computational earth sciences progress to even more challenging methods like artificial intelligence, it is time to change how we utilize existing methods and ideas to improve the reproducibility of our computational research. An investment in reproducible research design will pay off (Raphael *et al* 2020).

To progress towards more reproducible computational earth science that will ultimately foster trust in our model results, we need to move towards a holistic (manuscripts, data, and software development together in an Open Science framework under FAIR principles (Wilkinson *et al* 2016)) and transparent computational science (we propose such a path of change in figure S5). By utilizing Open Science guidelines and increasing our knowledge about tools, methods, and Open Source licenses, we can approach a future where sharing is the norm and where we can jointly verify and improve our research software.

With ever-increasing data amounts, e.g. through higher resolution satellite data and driven by the pressing matters of global change, earth and environmental research will remain at the frontier of big-data analysis and complex model development. To face these challenges, without losing sight of our scientific foundation and thus society's trust, we need to increase our efforts for reproducibility and adapt established code development methods that will increase code quality, code development efficiency, and finally, also increase reproducibility.

Data availability statement

Data, code, and instructions to replicate the analyses are available through the supplementary information and supplementary data.

Research. We furthermore thank the five reviewers that have helped to improve this manuscript.

Author contributions

R R led the conceptualization, formal analysis, visualization, and writing of the draft. Original idea was conceived by T T and R R. K S lead the methodology; T T, R R, and T W contributed to the design of the poll. T T, K S, and T W contributed to the initial draft regarding structure, wording, figures, and analysis.

Conflict of interest

No competing interests.

ORCID iDs

Robert Reinecke ⬤ https://orcid.org/0000-0001-5699-8584
Tim Trautmann ⬤ https://orcid.org/0000-0001-8652-6836
Thorsten Wagener ⬤ https://orcid.org/0000-0003-3881-5849

References

Añel J A, García-Rodríguez M and Rodeiro J 2021 Current status on the need for improved accessibility to climate models code *Geosci. Model Dev.* **14** 923–34

Baker M 2016 1500 scientists lift the lid on reproducibility *Nature* **533** 452–4

GMD 2019 Editorial: the publication of geoscientific model developments v1.2 *Geosci. Model Dev.* **12** 2215–25

Goodman S N, Fanelli D and Ioannidis J P A 2016 What does research reproducibility mean? *Sci. Trans. Med.* **8** 341–12

Hall C A, Saia S M, Popp A L, Dogulu N, Schymanski S J, Drost N, van Emmerik T and Hut R 2022 A hydrologist's guide to Open Science *Hydrol. Earth Syst. Sci.* **26** 647–64

Hut R W, van de Giesen N C and Drost N 2017 Comment on 'Most computational hydrology is not reproducible, so is it really science?' by Christopher Hutton *et al*: let hydrologists learn the latest computer science by working with research software engineers (RSEs) and not reinvent the waterwheel our *Water Resour. Res.* **53** 4524–6

Hutton C, Wagener T, Freer J, Han D, Duffy C and Arheimer B 2016 Most computational hydrology is not reproducible, so is it really science? *Water Resour. Res.* **52** 7548–55

Knowles R, Mateen B A and Yehudi Y 2021 We need to talk about the lack of investment in digital research infrastructure *Nat. Comput. Sci.* **1** 169–71

McConnell S 2016 *Code Complete: A Practical Handbook of Software Construction* 2nd edn (Redmond, WA: Microsoft Press)

Munafò M R, Nosek B A, Bishop D V M, Button K S, Chambers C D, Du Percie S N, Simonsohn U, Wagenmakers E J, Ware J J and Ioannidis J P A 2017 A manifesto for reproducible science *Nat. Hum. Behav.* **1** 0021

Nüst D, Sochat V, Marwick B, Eglen S J, Head T, Hirst T and Evans B D 2020 Ten simple rules for writing Dockerfiles for reproducible data science *PLoS Comput. Biol.* **16** e1008316

Peng R D 2011 Reproducible research in computational science *Science* **334** 1226–7

Perkel J M 2020 Challenge to scientists: does your ten-year-old code still run? *Nature* **584** 656–8

Raphael M P, Sheehan P E and Vora G J 2020 A controlled trial for reproducibility *Nature* **579** 190–2

Saltelli A *et al* 2020 Five ways to ensure that models serve society: a manifesto *Nature* **582** 482–4

Stagge J H, Rosenberg D E, Abdallah A M, Akbar H, Attallah N A and James R 2019 Assessing data availability and research reproducibility in hydrology and water resources *Sci. Data* **6** 190030

Stodden V, McNutt M, Bailey D H, Deelman E, Gil Y, Hanson B, Heroux M A, Ioannidis J P A and Taufer M 2016 Enhancing reproducibility for computational methods *Science* **354** 1240–1

Thornton P E, Cook R B, Braswell B H, Law B E, Corvallis W M, Shugart H H, Rhyne B T and Hook L A 2005 Archiving numerical models of biogeochemical dynamics *EOS Trans. Am. Geophys. Union* **86** 431

Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72

Wilkinson M *et al* 2016 The FAIR guiding principles for scientific data management and stewardship *Sci. Data* **3** 160018