Peer Review Report

Review Report on Reproducibility in subsurface geoscience

Original Research, Earth Sci. Syst. Soc.

Reviewer: Michiel van der Meulen
Submitted on: 06 Dec 2021
Article DOI: 10.3389/esss.2022.10051

EVALUATION

Q 1 Please summarize the main findings of the study.

The article discusses reproducibility in subsurface geoscience, based on a survey among more than 300 geoscientists from industry, academia, and government. The authors argue that that reproducibility is a significant issue, prime causes of which include data being inaccessible, conceptual frameworks being unclear or subjective, failure to quantify or report uncertainties, and methods being insufficiently described (including code being unavailable).

Q 2 Please highlight the limitations and strengths.

Concerning the scope: While the article suggests it covers reproducibility in subsurface geoscience in general, it mostly concerns reproducibility of such work as published in the peer-reviewed scientific literature. I think that this is too limited a scope (see Q.5).

Concerning the methodology: Even though I actually concur with the authors' main findings, I regret to say that the study does not comply with the publication standards they set or imply (see Q.5).

Q 3 Please comment on the methods, results and data interpretation. If there are any objective errors, or if the conclusions are not supported, you should detail your concerns.

The manuscript reads as an essay that happens to be supported by the results a questionnaire. I am not convinced that conducting a survey is the most appropriate way to assess the problem at hand (see Q.5).

Q 4 Check List

Is the English language of sufficient quality?
Yes.

Is the quality of the figures and tables satisfactory?
Yes.

Does the reference list cover the relevant literature adequately and in an unbiased manner?
No.

Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)
No.

If relevant, are the methods sufficiently documented to allow replication studies?
No.

Are the data underlying the study available in either the article, supplement, or deposited in a repository? (Sequence/expression data, protein/molecule characterizations, annotations, and taxonomy data are required to be deposited in public repositories prior to publication)
Yes.

Does the study adhere to ethical standards including ethics committee approval and consent procedure?
Not Applicable.

If relevant, have standard biosecurity and institutional safety procedures been adhered to?
Not Applicable.

Q 5 **Please provide your detailed review report to the editor and authors (including any comments on the Q4 Check List):**

SUMMARY
The article discusses reproducibility in subsurface geoscience, based on a survey among more than 300 geoscientists from industry, academia, and government. The authors argue that that reproducibility is a significant issue, prime causes of which include data being inaccessible, conceptual frameworks being unclear or subjective, failure to quantify or report uncertainties, and methods being insufficiently described (including code being unavailable). This results in a general disjoint between data and interpretations.

SCOPE AND OVERALL DESIGN OF THE RESEARCH
While the article suggests it covers reproducibility in subsurface geoscience in general, it mostly concerns reproducibility of such work as published in the peer-reviewed scientific literature. I think that this is too limited a scope. A lot, if not most of such research is undertaken outside academic contexts, for example to assess minerals, hydrocarbon or geological storage potential; to assess natural hazards; for general geological surveying; etc. The reproducibility of such work should be assessed on the supporting grey literature, typically bulky reports rather than the watered-down versions that are published in the peer-reviewed literature.
There is another scale aspect that the authors could have considered. Some research is not only (literally) too big to be comprehensively presented in a scientific journal, but also too big to reproduce. For example, the subsurface models I am personally responsible for are reproducible in the sense that the data input, conceptual framework, proprietary code, software used, etc. are all available (not only in the scientific literature, though). But who would have the resources to replicate a study that took 10s of FTE's worth of staff time to conduct? This implies that there should also be other mechanisms for other to ascertain the quality of modelling research than the option to replicate the results: most importantly an independent, transparent and auditable QC.
All in all, I am not convinced that the statement "publishing a paper in a peer-reviewed journal is considered the main output of a research project" is valid. I'd argue the opposite: "the scientific literature is inherently unsuitable to (exclusively) publish the results of large subsurface geoscience research projects."

METHODS
Even though I actually concur with the authors' main findings, I regret to say that their study does not comply with the publication standards they set or imply. The manuscript reads as an essay that happens to be supported by the results of a questionnaire. I am actually not convinced that conducting a survey is the most appropriate way to assess the problem at hand. The authors could for example have conducted a literature survey and grade reproducibility of a large number of studies on the basis of a checklist, or present some key case studies. They could even have stuck to writing an essay (without the questionnaire results). However, since they did choose to use a questionnaire, I will make a few comments on methodology.
The collection of qualitative scientific data by means of interviews or questionnaires is a proper scientific method (not in the geosciences, but for example in the social and medical sciences). The authors haven't consulted literature that discusses such methods, which would have inspired them to pay due attention to design the questionnaire in relation to the research goal. As result, they make some inferences that appear to be unsupported by the data they have collected. I'll give a few observations.

BIAS: The authors identified themselves as passionate about the problems and potential solutions of reproducibility within science. The high proportion of respondents who indicated that they tried to reproduce others' published (61%) suggests that the same may apply to the respondents. The authors highlight the problem of bias in subsurface geoscience research, but they haven't considered a bias in their own study. I am

not implying that passionate necessarily means overconcerned, but I do think that a group of concerned scientists will yield qualitative results that do not represent the community as a whole.

EXCLUSION AND WEIGHT OF DATA:
- 137 respondents (39%) indicated that they haven't tried to reproduce other's published results. So why didn't the authors discard their responses regarding the practical difficulties of reproducing published results, or at least weighed them differently?
- 41 respondents (12%) have the combination of a BSc or MSc degree and 0 to 5 years of experience. They could well be fresh graduates, who I'd say are not yet qualified to judge reproducibility in science.
- Beyond the a minimum experience level requirement, attention should have been made on the type of experience that the respondents have in subsurface related work. Beyond inferences that can be made from discipline, sector and industry, we need to at least be able to distinguish between the people who have hands-on experience with subsurface geoscience research and those who have not. In the first category I'd suggest to distinguish between for example project or quality managers (who oversee the whole work) and supporting staff (who don't, at least not necessarily). 17 respondents work in government, are the researchers or clients of researchers? Respondents in the latter category wouldn't necessarily have to be disqualified, but I think we need to know. Something similar goes for the two respondents who are unemployed and the one who is retired.
- The article discusses emphasizes reproducibility in the scientific literature, but 49 respondents indicated that have not been involved in scientific publishing (interestingly they seem to be somewhat more pessimistic about the reproducibility of subsurface geoscience than the ones who have. Why?)
- Being a post-doc researcher doesn't elevate your highest level of education above a PhD degree. It would have been more interesting to record whether the respondents have had post-academic training in aspects of subsurface research.

IN CONCLUSION
I think it is a good idea to raise concerns about reproducibility in subsurface geoscience research, and think of ways to raise the bar. I consider it a community issue rather than a subject-specific issue. After all, every peer-reviewed article reporting the results of such research without the means for others to reproduce it, is not only written by researchers who may have been driven by unhelpful incentives to publish scientifically. The work will have been reviewed by their peers and the decision to publish will have been made by the editor.
Even though I sympathize with the authors for raising the issue, my comments do not allow me to recommend anything but a rejection.

REGARDING THE SCORES UNDER QUALITY ASSESSMENT
- The scores for originality, rigor and overall quality of the study are consistent with the above remarks;
- I'll refrain from commenting on the quality of the writing (the authors are native speakers of English and I am not);

QUALITY ASSESSMENT

Q 6	Originality	
Q 7	Rigor	
Q 8	Significance to the field	
Q 9	Interest to a general audience	
Q 10	Quality of the writing	
Q 11	Overall quality of the study	

REVISION LEVEL

Q 12 What is the level of revision required based on your comments:

Substantial revisions.