The
Geological
Society

ES³ Earth Science,
Systems and
Society

Reproducibility in Subsurface Geoscience

Michael J. Steventon[1]*, Christopher A-L. Jackson[2], Matt Hall[3], Mark T. Ireland[4], Marcus Munafo[5] and Kathryn J. Roberts[6]

[1]Shell Research Ltd., Shell Centre, London, United Kingdom, [2]Department of Earth and Environmental Sciences, The University of Manchester, Manchester, United Kingdom, [3]Agile Scientific, Mahone Bay, NS, Canada, [4]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [5]School of Psychological Science, Bristol Population Health Science Institute, University of Bristol, Bristol, United Kingdom, [6]Institute for Global Health, University College London, London, United Kingdom

Reproducibility, the extent to which consistent results are obtained when an experiment or study is repeated, sits at the foundation of science. The aim of this process is to produce robust findings and knowledge, with reproducibility being the screening tool to benchmark how well we are implementing the scientific method. However, the re-examination of results from many disciplines has caused significant concern as to the reproducibility of published findings. This concern is well-founded—our ability to independently reproduce results build trust within the scientific community, between scientists and policy makers, and the general public. Within geoscience, discussions and practical frameworks for reproducibility are in their infancy, particularly in subsurface geoscience, an area where there are commonly significant uncertainties related to data (e.g., geographical coverage). Given the vital role of subsurface geoscience as part of sustainable development pathways and in achieving Net Zero, such as for carbon capture storage, mining, and natural hazard assessment, there is likely to be increased scrutiny on the reproducibility of geoscience results. We surveyed 346 Earth scientists from a broad section of academia, government, and industry to understand their experience and knowledge of reproducibility in the subsurface. More than 85% of respondents recognised there is a reproducibility problem in subsurface geoscience, with >90% of respondents viewing conceptual biases as having a major impact on the robustness of their findings and overall quality of their work. Access to data, undocumented methodologies, and confidentiality issues (e.g., use of proprietary data and methods) were identified as major barriers to reproducing published results. Overall, the survey results suggest a need for funding bodies, data providers, research groups, and publishers to build a framework and a set of minimum standards for increasing the reproducibility of, and political and public trust in, the results of subsurface studies.

Keywords: mining and SDGs, repeatability, replicability, reproducibility, CCUS, oil and gas, natural hazards

INTRODUCTION

Definitions of reproducibility can vary between and within disciplines. Here, we broadly define reproducibility as the ability to confirm the results and conclusions of your own or others' work. The principle can be split further into 1) repeatability, where results are obtained under the same conditions by the same research team, 2) replicability, where results are obtained by a different research team using the same methodology, and 3) reproducibility, where results are obtained by a different research team using a different methodology and/or dataset. Again, definitions can vary slightly between and within disciplines. The concept of reproducibility is vital to ensure complete reporting of all relevant aspects of scientific design, measurements, data, and analysis (see Goodman et al., 2016). In addition to these, transparency, which covers, access to data, software, documentation of methods, or metadata, for example, is a prerequisite of any study to facilitate repeatability, replicability, and reproducibility. Concern over the reproducibility of scientific results has gained significant traction in recent years as large-scale reviews in disciplines such as medicine (Nosek and Errington, 2017), psychology (Open Science Collaboration, 2015), and economics (Camerer et al., 2016), have often cast doubt on the reliability of publish results. A recent survey of scientists from across various fields found that >70% could not reproduce other scientists' experiments, and that >50% had failed to reproduce their own work (Baker 2016). This challenge in reproducibility has been attributed to several factors including poor method descriptions, selective reporting, poor self-replication, pressure to publish, poor/inadequate analyses, and lack of statistical power.

In 2015, researchers from a number of US academic and governmental institutions undertook a workshop focused on the theme of "Geoscience Paper of the Future" identifying that: 1) well-documented datasets hosted on public repositories, 2) documentation of software/code, including pre-processing of data and visualization steps and metadata, and 3) documentation and availability of computational provenance for each figure/result, would significantly aid efforts to improve reproducibility (see David et al., 2016). Here, we define subsurface geoscience as the study of Earth's material including rocks, gases and fluids not exposed at the surface. For the most part, subsurface geoscience has lacked any formal discussions or initiatives surrounding reproducibility. Part of this absence may be related to an emphasis on direct observations (rather than an experimentation focus), which often builds upon previous work (which itself has not been reproduced), creating a mixture of quantitative and qualitative data, and theoretical models (e.g., fault displacement models). These observations and models are integrated into broader studies, which include numerous case examples, and may be thought of, informally, as reproducibility studies combined to give a consensus.

Some areas of Earth science have been proactive in discussing and improving upon reproducibility, for example the reconstruct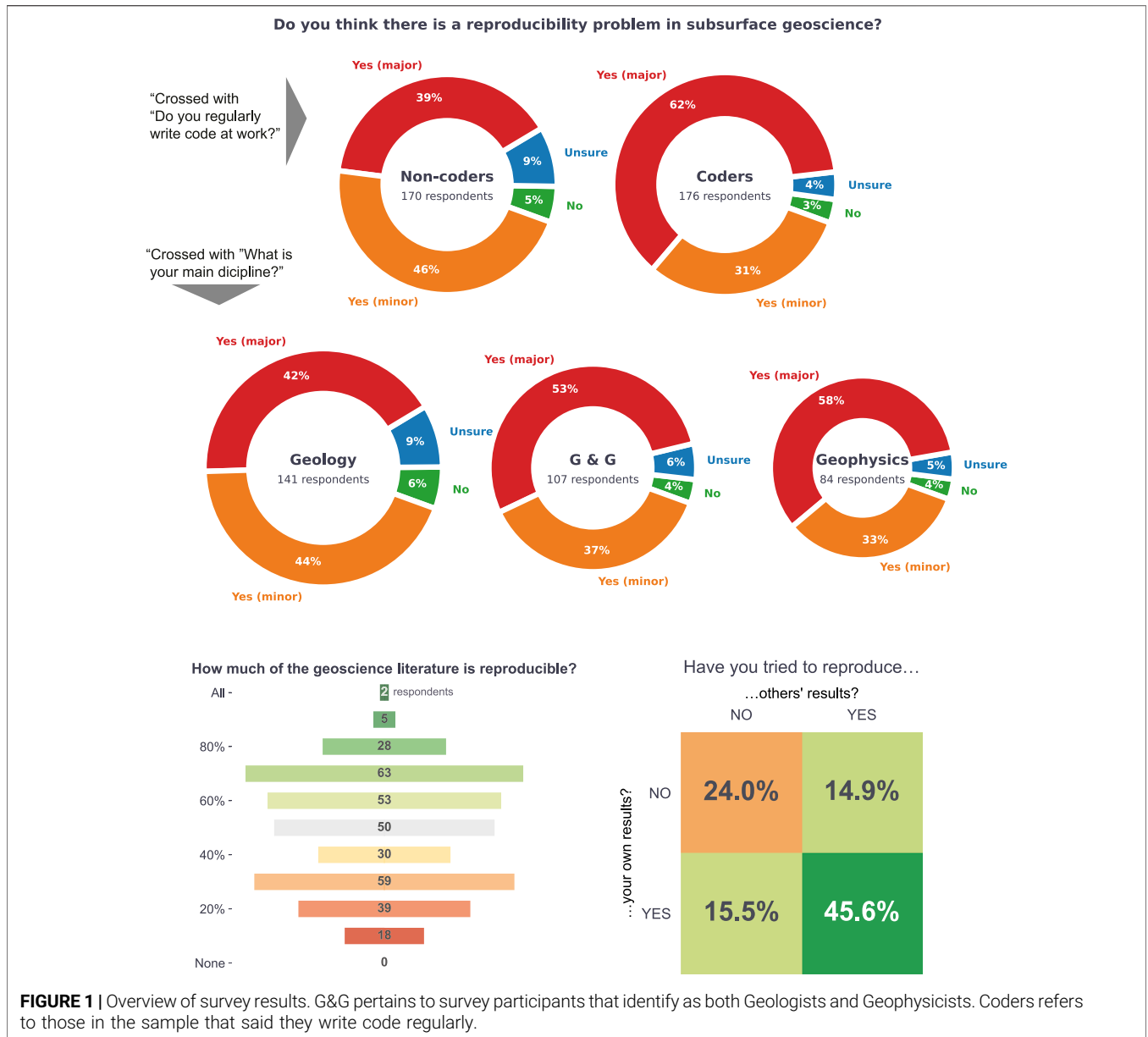ion of palaeo-climate through evaluation of atmospheric and sea-level changes (e.g., Milne et al., 2009). This field has been heavily scrutinised because of its direct impact on human lives, economies, and public policy when predicting and understanding climate change. Further examples of areas of geoscience addressing reproducibility include standardisation of pre-processing using open-source software in seismology (e.g., Beyreuther et al., 2010), large parts of computational geoscience (e.g., Konkol et al., 2019; Nüst and Pebesma, 2021), and reviews in quantitative areas of geomorphology (e.g., Paola et al., 2009; Church et al., 2020). Subsurface geoscience has received less attention with geoscientists often having little involvement with communication and outreach to policy makers, civic authorities, business leaders outside of related industries, media outlets, or the public at large (Stewart, 2016). However, in the future it is likely that areas in the subsurface such as carbon capture and storage (CCS), mining, and natural hazard assessment will receive increased levels of scrutiny, as their importance and interest by society grows.

Reproducibility and its modus operandi of transparency are critical to subsurface geoscience and should underpin the decision making for sustainable development and science communication. Transparency acts as a quality control mechanism which incentivises authors to publish results that are more robust and allows others to reproduce analyses. The aim of this study is to understand the current state of play of reproducibility in subsurface geoscience, collating geoscientists experience globally across academia, industry, and government.

SURVEY METHODS AND DATA

We use results collected from a 2020 survey of 346 geoscientists to assess views on the current state of reproducibility within subsurface geoscience. The survey utilised opportunity sampling (see Jupp et al., 2006) and was publicised through the author's networks, social media networks (e.g., Twitter), and geoscience-related mailing lists. Participants responded to 27 questions on topics including general questions on background, experiences, views on reproducibility, views on the practicalities of reproducibility and suggestions for improvements in subsurface geoscience going forward. Participants responded to closed questions (i.e., multiple choice) within the survey, with many providing extensive free-text comments.

The respondents were from a variety of backgrounds, including academia, governmental institutions, and multiple industries spanning mining & quarrying, engineering geology, oil and gas, renewable energy, hydrogeology, environmental monitoring, natural hazards, and CCS. The survey also achieved a global reach with respondents from the Americas, Asia, Africa, and Oceania, however, the majority were from Europe. The highest education level achieved by the participants ranged from undergraduate (7%), masters (28%), to doctoral (65%), and ranged in experience of

FIGURE 1 | Overview of survey results. G&G pertains to survey participants that identify as both Geologists and Geophysicists. Coders refers to those in the sample that said they write code regularly.

working with subsurface data from 0 to 20+ years. Most respondents identified themselves as geologists (41%), geophysicists (25%), or a combination of the two (30%). Over 85% of respondents said they publish formal literature. Just over half of respondents stated that they regularly write code at work.

The aim of this survey is to understand the views of the geoscience community in relation to reproducibility. The full survey presented to participants along with results and free-text comments can be found in the **Supplementary Material**. Results and associated comments have been anonymised, randomised, and redacted of any personal information to protect participants confidentiality. It must also be noted that all authors have a bias towards this issue, we are all passionate about improving reproducibility within science. In effect we have a theoretical bias towards this topic, and therefore, similarly to other surveys of this nature (e.g., Baker, 2016) the questions posed may have led respondents towards our biases.

RESULTS

We found that the majority (89%) of respondents identified the need for improved reproducibility in geoscience, with 38% stating there is a minor problem, and 51% stating there is a major problem (**Figure 1**). Broadly the responses were consistent across those who identified as geologists,
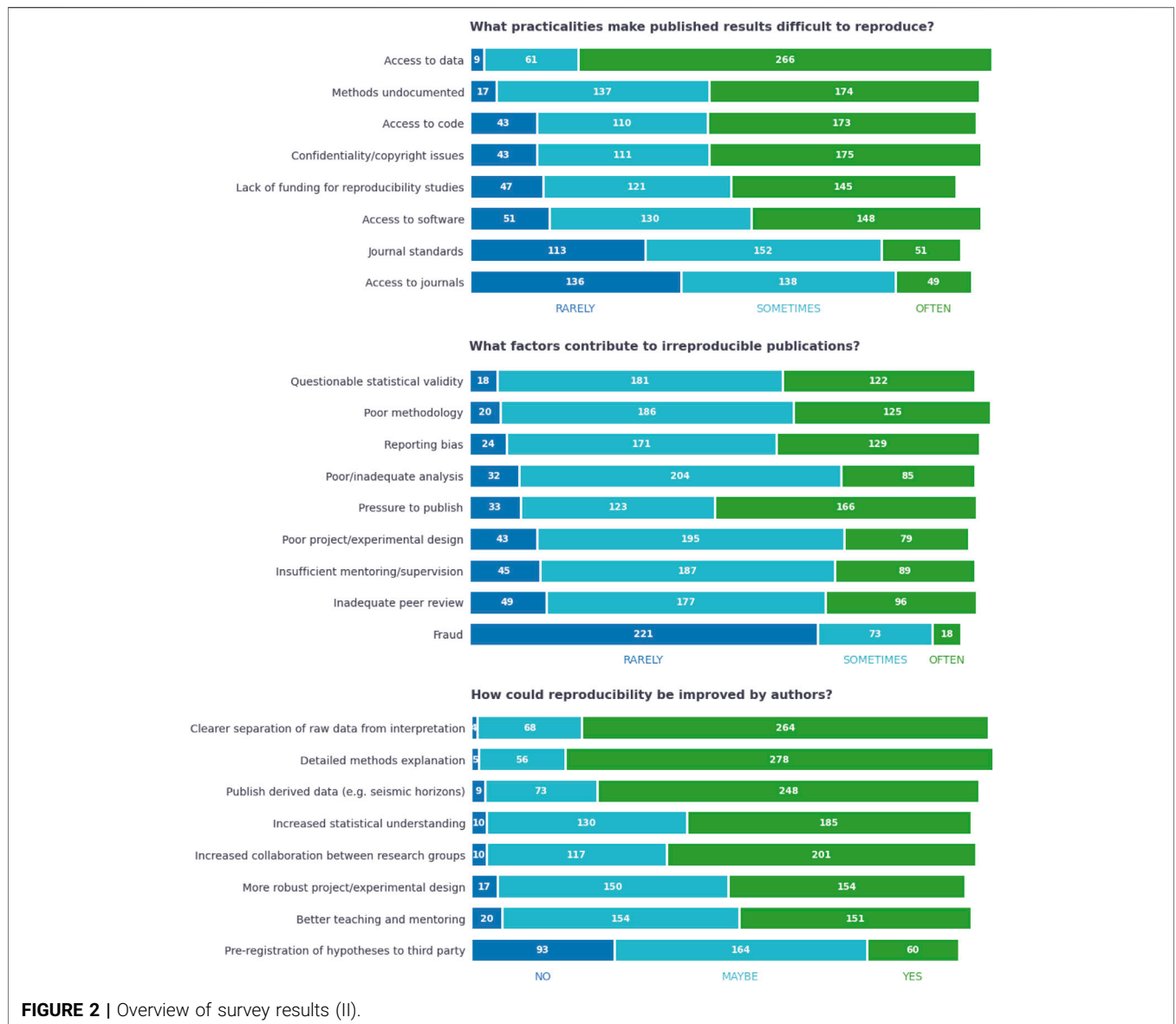
FIGURE 2 | Overview of survey results (II).

geophysicists, or integrated geology & geophysics professionals. Though when asked a follow up question about how much of the geoscience literature they considered is reproducible, we gathered unclear results with an approximately normal distribution of responses from 10% to 100%, with most respondents recording between 20% and 80% of the literature being reproducible (**Figure 1**). This suggests that although the community at large is aware there is a need to improve reproducibility, we are unsure as to how much of the literature is reproducible. Most noticeably those who write software code regularly seemed slightly more concerned/ aware about the problems of reproducibility (**Figure 1**).

Most respondents had tried to reproduce their own work at a later date, and 61% stated they had tried to reproduce others' published work. However, those that tried to reproduce both their own results and others results dropped to less than half. Reasons for reproducing an author's own work included:

carrying-out general quality control, testing different equipment, testing different methods and conceptual models, checking different software and code, and repeating analysis after the collection of additional data. However, only 13% of respondents have published or attempted to publish positive or negative replication findings, with many describing difficulties relating to data access, poorly described or incomplete methodology, an inability to access proprietary software, and inaccessible and/or poor data management. Behind these more practical issues, the survey revealed that people felt a lack of interest, hostility, and professional and/or personal cost to undertaking reproducibility studies, in particular for critiques of well-respected "seminal" papers. As one respondent noted "you are essentially critiquing work from a member of a very small close-knit community." In addition, those who have tried to publish reproducibility studies reported having been rejected at the peer-review

stage for reasons including editors and reviewers not considering replication studies as novel or ground-breaking science, not providing enough data from the original study, and conflicting interests with reviewers often having connections to the original study. In addition to this many comments highlighted concerns around conceptual/theoretical biases, in which prior experiences influence interpretations, and a lack of presenting alternative scenarios and models.

To better understand the barriers to improving reproducibility, we asked the participants what specific issues make published results difficult to reproduce. The most encountered difficulties included access to data (97%), undocumented methods (92%), confidentiality/copyright issues (85%), along with access to code (81%) or software (80%) (**Figure 2**). When asked what factors contribute to irreproducibility respondents identified several approximately equal issues including: reporting bias, poor methodology and analysis, questionable statistical validity, pressure to publish, poor project/experimental design, and insufficient mentoring and supervision of early-career scientists. Thankfully, fraud was viewed as the least likely factor to contribute to irreproducible publications. A number of reoccurring issues came up in the comments including: 1) a feeling that thorough peer-review is difficult to achieve as the reviewer typically only receives a mixture of finished figures and tables rather than the raw data in a usable and thus reproducible form, 2) the reduced time allotted to peer-review/the pressure to finish reviews on time, and 3) a pressure on authors to produce a larger quantity of research in less time, with the long-held mantras of "publish or perish" (e.g., Angell, 1986) and "quantity over quality" (e.g., Michalska-Smith and Allesina, 2017) still felt by many. These are known problems throughout many research fields, and it seems from this survey that these issues are also prevalent throughout geoscience.

DISCUSSION AND CONCLUSION

Our results are in line with other reviews of reproducibility (e.g., Baker 2016), and indicate that reproducibility is a significant issue in subsurface geoscience. Below we discuss some of the corrective measures that may aid in improving reproducibility, and challenges that persist.

Corrective Measures and Challenges
Data
Access to data was the most encountered difficulty identified by respondents (**Figure 2**). Within the free-text comments, many respondents detailed difficulties with reviewing papers, understanding and assessing results, and comparing studies. This is a well-known issue (e.g., Haibe-Kains et al., 2020) in subsurface geoscience, particularly when confidential industry data, for example from energy companies, is being used. Such data are not only costly to acquire, but are also commercially sensitive and therefore, cannot be openly shared. However, in some cases after a specified number of years these datasets may become openly available on

government repositories (e.g., National Data Repository [UK] and National Offshore Petroleum Information Management System [Australia]). One area where using open access data has proved useful for reproducibility is when testing machine learning algorithms on seismic-reflection surveys. Here use of open access data such as the popular Dutch F3 3D seismic-reflection survey (e.g., Waldeland and Solberg, 2017; Mosser et al., 2019) has allowed researchers to make direct comparisons. However, this approach is not suitable for every subfield of geoscience where a variety of different examples and data locations are required and therefore, different initiatives are needed to begin to bring some standardisation across varying datasets and localities.

The solution to the data access problem during the publication stage is in some ways simple, along with the publication itself, well-documented data sets on easy to access public repositories should be encouraged wherever possible. These shared data should also be granted Digital Object Identifiers (DOIs) so that they are indexed, and so credit is given to those compiling them. However, there are difficulties and complexities in trying to rectify this issue. Do we only publish work where data is open-access, or do we accept some studies using commercially sensitive data are useful to subsurface geoscience even if we cannot reproduce them? Some of the best quality data, in localities most publicly sponsored research could not fund, belongs to private enterprise. We could therefore, make critical, and possibly safety critical, observations and interpretations, and continue to garner knowledge from them. However, what's the value if there is no real clarity as to the robustness of the analysis, or if the results are not reproducible due to confidentiality in the methods employed? Where do we draw the line? This is a question for the community, but one potential solution we suggest would be to have some form of traffic light system on publications so readers can easily assess if the data is open or not. In addition, many subsurface dataset file sizes are large (i.e., 10-100s+ GB) hence data storage and management costs need to be taken into account.

Examples of good subsurface data sharing practices by governments include Geoscience Australia databases (e.g., NOPIMS) or the UKs Oil and Gas Authority National Data Repository. In addition, some private companies have also taken the initiative such Equinor's recent open-access distribution of subsurface data associated with the Volve oilfield, located on the Norwegian continental shelf. This dataset is relatively unique for academics as rather than been provided with a subsample of an industry dataset the entirety of the data can be used to allow a more comprehensive analysis of the subsurface. More initiatives like the above where industry data can be complied and shared openly will allow the democratisation of world class datasets and will hopefully increase reproducibility efforts.

Methods, Conceptual Bias, and a Framework for Minimum Publishing Standards
Geoscience is generally considered a derivative of the basic sciences (i.e., mathematics, physics, chemistry and biology.

Often geoscience involves a combination of experimental and interpretive (heuristics) methodologies. Frodeman (1995) reviews the methods employed in geological reasoning, where some aspects of geoscience are very much like the "basic sciences," objective, empirical, with certain and precise results (akin to physics), while other aspects are more interpretive and qualitative, where prior knowledge and experience provide a framework to "read" the Earth. The challenge of integrating these quantitative and qualitative results makes the Earth sciences both an intellectually stimulating discipline to study, but as we have seen also provides a challenge for implementing reproducibility.

Separation of interpretations from raw data and/or observation, access to code, conceptual uncertainty, along with lack of a detailed methodology framework repeatedly ranked as major issues in the survey (**Figure 2**). Subsurface geoscience benefits from a breadth of geological, geophysical, biological, and chemical data and associated methodologies. As a result of this the methods, software and code used in the analysis of data span many different disciplines, and similar to Bond et al. (2007) our survey also highlights the importance of conceptual bias/ uncertainty and the impact this may be having on reproducibility. The results of this and previous studies suggest that: 1) uncertainty should be quantitatively or qualitatively captured in subsurface geoscience studies, where data quality and/or limitations in current knowledge mean there is no definitive conclusion (e.g., Frodeman, 1995; Alcalde et al., 2017), 2) methods should be encouraged to standardise measurements (where possible) so studies can be compared on a like for like basis (e.g., Clare et al., 2019), and 3) methods should promote the documenting of alternate scenarios (e.g., Bentley and Smith, 2008; Bond et al., 2008). Clare et al. (2019) demonstrates how a consistent global approach to the measurement of subaqueous landslides enables comparison of research across scales and geological environments and has been adopted by much of the research community and governmental organisations. Bentley and Smith (2008) showcase how modelling subsurface reservoirs using multiple deterministic realisations, rather than anchoring on one realisation, increases the chance of capturing the full uncertainty range within a hydrocarbon or CCS development project. The above examples suggest a set of guidelines for publications focusing on detailed documentation of methods and conceptual uncertainty would be beneficial to enhancing reproducibility efforts. This does not necessarily mean lengthening the paper, but instead providing adequate and detailed **Supplementary Material** for readers who are aiming to reproduce work. We therefore draw similar conclusions to the "Geoscience Paper of the Future" (see David et al., 2016), calling for further transparency of methods employed. Ironically many publishers have strict standards in place for type setting and formatting of manuscripts, while often neglecting the more important aspects of the actual science. Hence, we would echo the thoughts of David et al.

(2016) on the matter of implementing some form of minimum standards framework for published work.

Although not formally considered in the context of this study grey literature (i.e., research and technical work produced outside of academic publishing and distribution channels) contributes a significant amount to subsurface geoscience knowledge. Often studies within industry have access to more comprehensive datasets, proprietary technologies, and greater resources than academics and governmental organisations. In most cases study results stay within a company or organisation, principally due to commercial value, the need for confidentiality, and a lack of incentives from a commercial perspective to publish. In addition, there are also challenges similar to large research grants in comprehensively presenting multi-year or even decadal projects where outputs generated from multiple scientists and groups do not concisely fit into the standard research publication format. Although much of this work stays confidential a lot can be learnt from best practices in some industry-settings where auditable data inputs, quality control and assurance, internal and external peer-assists, tested workflows and processes, and re-analysis of previous data are commonplace—essentially a business focused system following many of the requirements for reproducibility.

Incentives

Currently publishing a paper in a peer-reviewer journal is considered one of the main outputs of a research project for academics. This is because journal publication underpins citations metrics such as the h-index, which are commonly used to assess research outputs and researchers. The more papers a researcher publishes and the more these papers are cited, the greater the esteem and the higher the likelihood they will be hired, promoted, or rewarded. As such, from a metrics standpoint it is more important to publish cited papers, regardless of the quality and, certainly, the reproducibility of the work. Another citation metric is the Journal Impact Factor or "JIF," which essentially attempts to communicate the quality of a journal and thus, the quality of the papers published within that journal. Many general (e.g., Nature, Science) or field-specific (e.g., Geology, Nature Geoscience), high-JIF journals are highly prized by scientists for the reasons outlined above; however, the conflict here is that these journals typically favour papers containing "novel" or "ground-breaking" work of general interest, with studies focused on reproducibility or replication less likely to be published. There is also less value attached to publishing so-called negative results, despite the underpinning studies being central to refining and improving the scientific method and assessing the robustness (and in some cases safety) of already published work.

The knowledge, adoption, and quality of reproducibility in subsurface geoscience could thus be increased by modifying current incentive frameworks in academic and non-academic workplaces. Greater emphasis and value

needs placing on research methods and transparency, and the overall robustness of results, as at least partly defined by how reproducible they are. This could be done by incentivising open and transparent science, the basis for reproducibility, in hiring, promotion, and general assessment frameworks. For example, credit (in whichever form is appropriate) should be given for committing to providing datasets and derivative data (e.g., seismic interpretation horizons), code, and software, that is publicly available. This can be achieved by placing these data in public repositories that issue DOIs. Currently there is a political drive for open research by many of the global funding bodies (e.g., UKRI, DFG, NSF) who require data and outputs to be publicly available. In addition to this, organisations such as UK Reproducibility Network are supporting and promoting reproducibility within multiple scientific disciplines (see **Box 1**).

Future Work

To begin to assess and quantify how much of the subsurface geoscience literature is reproducible we propose further investigations using a systematic review-based approach to grade the reproducibility of sub-topics within the geosciences. Similar studies have been undertaken in geospatial research such as Konkol et al. (2019) who demonstrated technical issues with spatial statistics results they reproduced from several papers. In addition to these, targeted studies focusing on specific type localities which inform many of our subsurface models, such as Madof et al. (2019) reassessment of the Mississippi Fan sequences should be encouraged.

Reflections

Reproducibility is an essential element of scientific work. It enables researchers to re-run and re-use experiments/studies reported by others, learning from their successes and failures, and by doing so producing overall "better" science. Geosciences has a significant contribution to make in implementing the United Nations (UN) Sustainable Development Goals (SDGs) through collaborating with social scientists, partnering with civil society and end users, and communicating existing research to policymakers (Scown, 2020). Subsurface geoscience is core to this, where the work has a significant role in informing and supporting the delivery of projects such as CCS, nuclear waste disposal, mining for base metals and rare Earth elements, and forecasting relating to the mitigation of natural hazards (e.g., volcanic eruptions and earthquakes). In particular, those activities, which involve geoengineering, are often seen as controversial, in their infancy in terms deployment or understanding (e.g., Shepherd, 2009), and capital-intensive requiring oversight from governmental and intergovernmental bodies. Improved reproducibility and transparency will allow the weight of evidence presented to be evaluated more efficiently and reliably, allowing the design of a higher proportion of future studies to address actual knowledge gaps or to effectively strengthen cumulative evidence (Scown, 2020).

Today there exists the digital infrastructure to support the publishing and archiving of results and data to enable geoscientists to carry out and assess reproducibility. There will always be hurdles to overcome however, we believe it is important for the geoscience community to embrace the importance of reproducibility, particularly as our science will likely be put under the spotlight in the coming decades. Our results also highlight the notion of reproducibility and the requirement for substantive and trustworthy data beyond geoscience to disciplines of natural and social sciences. We hope the survey results provide an initial catalyst for conversations within academia, industry, and government, around how we can begin to improve working practices and dissemination of geoscience to specialists, practitioners, policy makers, and the public.

BOX 1 | A whole-system approach to improving reproducibility.

Scientific research is a human endeavour, and therefore fallible, subject to cognitive biases brought by scientists themselves to their work, and shaped by incentives that influence how scientists behave (Munafò et al., 2020). Understanding the nature of the research ecosystem, and the role played by individual researchers, institutions, funders, publishers, learned societies and other sectoral organisations is key to improving the quality and robustness of the research we produce (Munafò et al., 2017). Coordinating the efforts of all of these agents will be necessary to ensure incentives are aligned and promote appropriate behaviours that will drive research quality, and reward the varied contributions that diverse individuals make to modern scientific research.

The UK Reproducibility Network (UKRN; www.ukrn.org) was established to support the coordination of these agents. It comprises local networks of researchers, institutions that have formally joined the network, and external stakeholders comprising funders, publishers, learned societies and a range of other organisations. It works to coordinate activity within and between these groups, with the overarching goal of identifying and implementing approaches that serve to improve research quality. Rather than treating issues such as the low reproducibility of published research findings as a static problem to be solved, it recognises the dynamic and evolving nature of research practices (and research culture more generally), and in turn the need to adopt a model of continual improvement.

One area of current focus is open research—and transparency in research more broadly—recognised by the UK government R&D People and Culture Strategy as "integral to a healthy research culture and environment." This allows the recognition of more granular, intermediate research outputs (e.g., data, code) from diverse contributors to the research process. However, it has also been argued to serve as a quality control process, allowing external scrutiny of those outputs, and in turn creating an incentive to ensure robust internal quality assurance processes (Munafò et al., 2014). However, fully embedding open research practices will require infrastructure and training, incentives (for example, recognition in hiring and promotion practices), support from funders and publishers, and more.

We also need to recognise that well-intentioned changes to working practices may not work as intended or, worse still, may have unintended consequences. For example, they may serve to exacerbate existing inequities if only well-resourced research groups are able to engage with them fully, particularly if this engagement becomes a factor that influences grant and publishing success. We therefore need to evaluate the likely impact of changes to working practices on under-represented and minoritised groups, and continue to evaluate the impact of these changes once implement. Again, we need to move to a model of continual improvement which links innovation to evaluation. Meta-research—also known as research on research—is now an established area of inquiry that allows us to use scientific tools to understand (and improve) the process of science itself. Supporting this activity will be central to successful, positive change.

DATA AVAILABILITY STATEMENT

The full survey presented to participants along with results and free-text comments can be found in the **Supplementary Material**. Results and associated comments have been anonymised, randomised, and redacted of any personal information to protect participants confidentiality.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

CONFLICT OF INTEREST

Author MS was employed by the company Shell Research Ltd. Author MH was employed by the company Agile Scientific.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.escubed.org/articles/10.3389/esss.2022.10051/full#supplementary-material

REFERENCES

Alcalde, J., Bond, C. E., Johnson, G., Ellis, J. F., and Butler, R. W. (2017). Impact of Seismic Image Quality on Fault Interpretation Uncertainty. *GSA Today* 27, 4–10. doi:10.1130/gsatg282a.1

Angell, M. (1986). Publish or Perish: a Proposal. *Ann. Intern. Med.* 104 (2), 261–262. doi:10.7326/0003-4819-104-2-261

Baker, M. (2016). Reproducibility Crisis. *Nature* 533 (26), 353–366. doi:10.1038/533452a

Bentley, M., and Smith, S. (2008). Scenario-based Reservoir Modelling: the Need for More Determinism and Less Anchoring. *Geol. Soc. Lond. Spec. Publ.* 309 (1), 145–159. doi:10.1144/sp309.11

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. (2010). ObsPy: A Python Toolbox for Seismology. *Seismol. Res. Lett.* 81 (3), 530–533. doi:10.1785/gssrl.81.3.530

Bond, C. E., Gibbs, A. D., Shipton, Z. K., and Jones, S. (2007). What Do You Think This is? "Conceptual Uncertainty" in Geoscience Interpretation. *GSA today* 17 (11), 4. doi:10.1130/gsat01711a.1

Bond, C. E., Shipton, Z. K., Gibbs, A. D., and Jones, S. (2008). Structural Models: Optimizing Risk Analysis by Understanding Conceptual Uncertainty. *First Break* 26 (6). doi:10.3997/1365-2397.2008006

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., et al. (2016). Evaluating Replicability of Laboratory Experiments in Economics. *Science* 351 (6280), 1433–1436. doi:10.1126/science.aaf0918

Church, M., Dudill, A., Venditti, J. G., and Frey, P. (2020). Are Results in Geomorphology Reproducible? *J. Geophys. Res. Earth Surf.* 125 (8), e2020JF005553. doi:10.1029/2020jf005553

Clare, M., Chaytor, J., Dabson, O., Gamboa, D., Georgiopoulou, A., Eady, H., et al. (2019). A Consistent Global Approach for the Morphometric Characterization of Subaqueous Landslides. *Geol. Soc. Lond. Spec. Publ.* 477 (1), 455–477. doi:10.1144/sp477.15

David, C. H., Gil, Y., Duffy, C. J., Peckham, S. D., and Venayagamoorthy, S. K. (2016). An Introduction to the Special Issue on Geoscience Papers of the Future. *Earth Space Sci.* 3 (10), 441–444. doi:10.1002/2016ea000201

Frodeman, R. (1995). Geological Reasoning: Geology as an Interpretive and Historical Science. *Geol. Soc. Am. Bull.* 107 (8), 960–0968. doi:10.1130/0016-7606(1995)107<0960:grgaai>2.3.co;2

Goodman, S. N., Fanelli, D., and Ioannidis, J. P. (2016). What Does Research Reproducibility Mean? *Sci. Transl. Med.* 8 (341), 341ps12. doi:10.1126/scitranslmed.aaf5027

Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Waldron, L., Wang, B., et al. (2020). Transparency and Reproducibility in Artificial Intelligence. *Nature* 586 (7829), E14–E16. doi:10.1038/s41586-020-2766-y

Jupp, V. (2006). *The SAGE Dictionary of Social Research Methods*, Vols. 1–0. London: SAGE Publications.

Konkol, M., Kray, C., and Pfeiffer, M. (2019). Computational Reproducibility in Geoscientific Papers: Insights from a Series of Studies with Geoscientists and a Reproduction Study. *Int. J. Geogr. Inf. Sci.* 33 (2), 408–429. doi:10.1080/13658816.2018.1508687

Madof, A. S., Harris, A. D., Baumgardner, S. E., Sadler, P. M., Laugier, F. J., and Christie-Blick, N. (2019). Stratigraphic Aliasing and the Transient Nature of Deep-Water Depositional Sequences: Revisiting the Mississippi Fan. *Geology* 47 (6), 545–549. doi:10.1130/G46159.1

Michalska-Smith, M. J., and Allesina, S. (2017). And, Not or: Quality, Quantity in Scientific Publishing. *PLoS one* 12 (6), e0178074. doi:10.1371/journal.pone.0178074

Milne, G. A., Gehrels, W. R., Hughes, C. W., and Tamisiea, M. E. (2009). Identifying the Causes of Sea-Level Change. *Nat. Geosci.* 2 (7), 471–478. doi:10.1038/ngeo544

Mosser, L., Oliveira, R., and Steventon, M. (2019). "Probabilistic Seismic Interpretation Using Bayesian Neural Networks," in 81st EAGE Conference and Exhibition 2019, London (London: European Association of Geoscientists & Engineers), 1–5.

Munafò, M., Noble, S., Browne, W. J., Brunner, D., Button, K., Ferreira, J., et al. (2014). Scientific Rigor and the Art of Motorcycle Maintenance. *Nat. Biotechnol.* 32 (9), 871–873. doi:10.1038/nbt.3004

Munafò, M. R., Chambers, C. D., Collins, A. M., Fortunato, L., and Macleod, M. R. (2020). Research Culture and Reproducibility. *Trends Cognitive Sci.* 24 (2), 91–93. doi:10.1016/j.tics.2019.12.002

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., et al. (2017). A Manifesto for Reproducible

Science. *Nat. Hum. Behav.* 1 (1), 0021–0029. doi:10.1038/s41562-016-0021

Nosek, B. A., and Errington, T. M. (2017). Reproducibility in Cancer Biology: Making Sense of Replications. *Elife* 6, e23383. doi:10.7554/elife.23383

Nüst, D., and Pebesma, E. (2021). Practical Reproducibility in Geography and Geosciences. *Ann. Assoc. Am. Geogr.* 111 (5), 1300–1310. doi:10.1080/24694452.2020.1806028

Open Science Collaboration (2015). Psychology Estimating the Reproducibility of Psychological Science. *Science* 349 (6251), aac4716. doi:10.1126/science.aac4716

Paola, C., Straub, K., Mohrig, D., and Reinhardt, L. (2009). The "Unreasonable Effectiveness" of Stratigraphic and Geomorphic Experiments. *Earth-Science Rev.* 97 (1-4), 1–43. doi:10.1016/j.earscirev.2009.05.003

Scown, M. W. (2020). The Sustainable Development Goals Need Geoscience. *Nat. Geosci.* 13 (11), 714–715. doi:10.1038/s41561-020-00652-6

Shepherd, J. G. (2009). *Geoengineering the Climate: Science, Governance and Uncertainty*. London: Royal Society.

Stewart, I. (2016). Sustainable Geoscience. *Nat. Geosci.* 9, 262. doi:10.1038/ngeo2678

Waldeland, A. U., and Solberg, A. H. S. S. (2017). "Salt Classification Using Deep Learning," in 79th eage conference and exhibition 2017, Paris (Paris: European Association of Geoscientists & Engineers), 1–5.