**Reviewer: 1**

The study aims to increase our understanding of how researchers in the geosciences view replication, using a sampling frame that breaks with the past convenience samples. This is interesting and relevant work worth publishing about. However, I feel that the current version of the manuscript does not interpret and exploit the survey outcomes to the maximum extent possible. Below some suggestions for improvements:

*Response and Organizational Note:* *We thank this reviewer for their careful consideration of our paper and these very helpful comments. We engage with each comment below bearing in mind the point to take further advantage of our survey outcomes. As an organizational note, we have slightly rearranged the reviewer comments below by grouping comments that addressed similar issues. All 13 original comments are preserved, and our responses are identified by line here and highlighted in red in the revised text.*

1. Overall, the background contextualization and the conceptualization of important elements are convincingly developed. Nevertheless I struggle to understand the place and role of the concept of direct replication (p10f). What is the difference to reproduction and conceptual replication, and more importantly, does it matter for the purposes of this paper? It is not being taken up again in the discussion.
2. How does the survey's definition of replication relate to the more detailed one from the Turing Way?
3. The nomothetic-idiographic debate goes decades back, and it might be worth mentioned at least some of the original papers by Hartshorne and Schaefer here (p6).

*Response: We have responded to each of these comment in the ‘Replication and the Evaluation of Prior Claims’ section of the paper. The first paragraph of this section now includes a discussion of definitions of replication that draws from the Turing Way and he broader literature. We have also added direct references to the original Hartshorne Schaefer debate and the early literature it spawned. Finally, while we agree with the reviewer that the distinction between direct and conceptual replication is not used extensively in later portions of the paper, we have retained these terms to maintain connection to the broader literature where their use is somewhat common. However, we have also worked to make clearer connections to these concepts in our revised discussion and conclusions section.*

1. In the sampling scheme, what is the justification for the JIF of 1.5? Even within the geosciences, JIF vary widely between sub-disciplines, and there is enough evidence that the JIF often does not reflect well a journal's scientific impact or merit.
2. While I am not trying to advocate statistical hypothesis testing where it's ill-advised because of small sample sizes, I was wondering whether some questions could be addressed through more than comparing and interpreting percentages.

*Response:* We selected JIF of 1.5 as a practical stopping rule for data collection. We agree that there are contributions in the literature outside of this group of journals. However, our goal was to develop a robust list of authors for our survey. We believe that sampling from these journals over the 5-year period gives us a fairly exhaustive list that is representative of scholars active in the field. We also gave extensive thought to performing formal statistical tests. However, we concluded that our sample sizes and variation in how researchers appeared to conceptualize key ideas would undercut the validity of those formal comparisons. Instead, we provide a through description of directional differences and identify variation in the magnitude of selective responses. Moreover, all of our procedures and raw data are fully available, which allows any researcher to explore data patterns as they see fit. We believe this is a more complete and justified way to communicate our results, as compared to somewhat shaky statistical testing.

1. It seems that most researchers do not know what is necessary for a successful replication (p20). This is itself and important outcome, showing that as a discipline we might have to continue the debate on what types of outcomes and insights we aim for.
2. Table 4 shows important information, but it is not straightforward to link the categories to the questions in the annex, because they are diffently formulated. E.g., the question in the survey is "How often does the low value of replication studies affect researcher's decisions to attempt a replication study in your sub-field". That is not quite the same as the table, which shows "How often does the perceived value affect researcher's decisions". The survey question is much more suggestive than the table shows. It's also not quite clear how researchers understood the "value" here - for their career? For science? The respondents seem to think that much geographic research is worth replicating (figure 2) but later arguments disagree (see #12).
3. An interesting bit of information would be how many respondents initially considered a replication but then decided early (before actually starting) that they not continue, and for which reason they stopped. At the moment, we only know how many actually attempted a replication, not how many considered it.
4. On the reasons for not publishing the replication results, it would be interesting to hear whether there were any respondents who experienced rejections because of a perceived lack of originality, as this was voiced initially as one of the main reasons against replication studies.

*Response:* We agree with the observation of (6) and added interpretation and implications of this to the discussion. We have addressed the concern raised in (7) by presenting the survey question in text and explaining its connection to Figure 4. We have also adjusted the category labels in selected to more clearly communicate the survey question. Lastly, we added a few sentences on the interpretation of value. Text incorporating the comments 6, 8, 9 have been added to the new discussion and conclusions sections. We added some insight into (8) from qualitative responses about replications that were started but not published. Although this sample of respondents "gave up" earlier than the comment inquires about, we can reasonably assume that researchers abandoned the effort for similar reasons.

1. Footnote 1 seems to contain some mistake.
2. There are occasional formatting problems with references, e.g., p12 L5
3. Not sure I understand the argument given by respondents and mentioned on p26 L20f - how can a geospatial study lack information on where it was conducted?
4. An interesting interpretation of the responses is that many studies do not contain findings worth reporting (and thus not worth validating through replication), compare p28 L28.

*Response:* We have made each of these adjustments in the manuscript. Related to point 12, we have added a sentence contextualizing the result in reference to the findings of prior studies. While surprising, it seems that many published studies do not share sufficient details about the precise location of spatial extent or their work (e.g., a geographically referenced bounding box, or an exact Landsat tile reference) see Margulies et al. (2016) in the *Annals* and Aabeyir 2023 in *African Geographical Review*.

**Reviewer: 2**

Thank you for the opportunity to read this paper. The study investigates in a scientifically sound way an important topic that is often ignored in the discussion about R&R, this of replicability. The sampling method for the conducted survey is a strong point of the work, and it belongs to the most convincing ones I have seen so far in studies about the perceptions of R&R within the scientific community.

The structure of the work is overall good, with the exception of the final section, named "Discussion and Conclusion". Although I cannot be sure, since the numbering of the sections and subsections is missing in the document that I received, I assume that "Limitations and Future Work" is also part of this section. The latter reads more as a conclusion rather than describing limitations. I would recommend to separate the Discussion (with limitations) and the Conclusion in two different sections to conclude this work concisely.

*Response:* *We thank the reviewer for their supportive assessment and suggestions. Following the suggestion, we have separated the Discussion and Conclusions into distinct sections with the limitations placed within the discussion.*

The document begins (p. 3, l. 2-9) with definitions for R&R. I would recommend to provide citations for these definitions to avoid confusion, as there are two major approaches regarding what is replicability and what is reproducibility (for reference, see <https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html#rr-overview-definitions>, <https://arxiv.org/abs/1802.03311>). Both of these "schools" are widely adopted, so I think it is advantageous for this specific paper to acknowledge the existence of both and to clarify which one is adopted by the authors. This might also be helpful in the discussion section (p. 27, l. 18-19), where the authors note that researchers tend to conflate the two terms (there is a good reason for this).

The term "epistemological functions" is a recurring one throughout the document. Figure 1 provides some sort of indication about what is meant with this term, but I think it will enhance the clarity of the manuscript to define the term in the text and to provide some examples of epistemological functions.

*Response: we have expanded our discussion of the definition of replication and their use across disciplines in the ‘Replication and the Evaluation of Prior Claims’ section of the paper. The first paragraph of this section now includes a discussion of definitions of replication that draws from the Turing Way and he broader literature. To further locate the definitions, we have added a few sentences and references to a series of papers/reports that map usage of these terms in different disciplines. Following the reviewer’s suggestion, we have also highlighted the connection between definitional conflation and varying use of terms across disciplines in the discussion. We have replaced "epistemological function" with the simpler and more direct "epistemic purpose" in all instances of the paper and given some specific examples.*

More comments that could enhance the clarity of the document can be found below (many of them just refer to syntax and typos):

**Introduction**

* p. 5, l. 6-9: It is not clear to me what is the implication of this sentence.
* p. 5, l. 25-29 and p. 6, l. 1-10: Is this a taxonomy from the literature or is it introduced by the authors? This should be clarified.

*Response:* *We have reworked each of these sections to improve clarify. The taxonomy we present is a synthesis of taxonomies presented by the prior authors cited in the draft. However, we have reworked the writing to make the connection and attribution clearer as suggested.*

**The Replicability of Geographic Research**

* p.8, line 29: small typo; argue rather than argues?
* p. 9, line 4: "case of location" reads a bit confusing.
* p. 9, lines 19-21: "Moreover... location": this sentence seems quite general. Why or in which cases it is unclear which factors should be accounted for when studying a location? Or give examples of such factors maybe

*Response: We have corrected and added details that address these issues. We have separated factors related to study design from factors intrinsic to locations and added an example.*

**Data and Methods**

* p. 9, l. 8-16: OSF repositories can be provided as anonymized links for double-blind peer-review purposes.
* p. 12, l. 5: nosek2020: typo on a citation?
* p. 13, l. 14: What is meant by "research compendium" in this case?

*Response:* W*e have corrected the citation typo and changed "compendium" to "repository" for consistency with our language at the beginning of this section. More broadly, the note on the anonymized link is a very useful. We have used Git version control and GitHub to transparently track the full project life cycle and to publicly register intermediary research products. Our OSF project integrates seamlessly with the GitHub repository, making impossible to remove identifying metadata. However, this has given us something to think about for our next project. Perhaps in future we will create an OSF repository with the essentials for anonymous review and upon publication, link and populate that repository back with all of our version tracked work from GitHub. Simpler still, we have attempted to share all our links and work with our names attached at the start of review, but our offers to waive our right to blinding have not yet found purchase at different journals.*

**Results**

* p. 14, l. 22-24: What are the different career levels? Does it refer to the question "which job title best applies to you" of the questionnaire?
* p.16, l. 2: What makes the definitions "interpretable"?
* p. 17, l. 17: Syntax is a bit confusing

*Response: We have altered the text to address these comments. p14 does refer to job title. We changed the wording in the text to better capture this idea. We have dropped interpretable. There are a small number of "NA" or similar text responses that were user provided, but are not definitions of the term. However, nearly all provided definitions could be read, so the adjective is not necessary and confusing as suggested.*

* Figure 3: "researcher position": Is it meant as in rank or special mission, eg. in Antarctica? This is not specified in the questionnaire as well.
* p. 19, l. 9-13: According to Figure 3, there are certain tendencies in the participants' responses, while the text implies that they are quite evenly distributed.
* p. 20, l. 17: across all characteristics except for "Not measurable" (only for the "strongly disagree"). Overall all I agree with the bottomline of this paragraph, with this exception, which is showing a clearer tendency compared to the other five phenomenon characteristics.
* p. 20, l. 20-21: Spatial dependence has just a stronger uncertainty (NE/DK). "Likely to decrease" sums also to 41% for spatial variability (same as spatial dependence).
* p. 20, l. 25-26: "respondents were split" - they have a difference of 8%. It's a borderline case.
* p. 21, l. 5-6: "with no clear trend": again, the difference between positive and negative is 11%, so I am wondering how the authors decide if the the difference is big enough to show a trend or not.

*Response:* T*hank you for this series of carefully considered and helpful comments. We have revised each of these text sections improve our presentation and interpretation of these results. We have adjusted our language on several points in an attempt to not overstate, or over interpret our data. Our new presentation is much more cautious and places the directional indicators around each characteristic in the context of the 'don't know' and 'no effect' results. We believe our new writing conveys an indication of direction coupled with overall uncertainty among the respondents.*

*As a clarification, by researcher position, we originally intended to capture the concept of positionality, which is an important element of qualitative researcher design. However, it is certainly possible that a survey respondent might also read the question as the official position, or even location of a researcher. To address this concern, we have added some additional writing to the new discussion section to reflect the ambiguity of subfield-specific terminology like "researcher position" and its implications for debates about replication in the field.*

* p. 21, l. 12: Is this restricted only to a specific subfield ("geographic" being used as a descriptor of the sample)
* p.21, l. 34: difficulty in accessing
* p. 22, l. 8-9: "or the belief that replications have high chances of success": Seems a bit contradicting with the response analysed, can you please clarify?
* Table 1 and Figure 4 seem to have the same groupings on the vertical axis with different labels (eg., Research Environment vs Academic Incentives). It is less confusing if the labeling becomes consistent for both.
* p. 25, l. 13: "Before" does not stem from the questionnaire, as it only asks if it was done in both locations without specifying the order

*Response:* *We have made and tracked adjustments to each of these comments in the text. Thank you for catching the inconsistency between Fig 4 and Table 1, and our inclusion of 'before' when it was indeed not in the question. We have corrected both these issues.*

**Discussion:**

* I think that possible overlaps between the different subfields of the respondents should also be addressed here. For example, researchers that study topics related to climate change, environmental policies, natural resource management, etc. could identify with both "environmental sciences" and "nature and society" subfields.
* p. 27, l. 20-22: That's an excellent comment.
* p. 29, l. 1: Citation error?
* p. 29, l. 6: researcher -> researchers?
* p. 31, l. 18: one -> in?

*Response: We have included in the revised discussion commentary on the potential overlap between researcher subfields and the implications of that overlap on researcher response. We have also made these remaining minor edits in the manuscript.*