Author Response to Referee Comments - egusphere-2026-491

We sincerely thank the Editor and the reviewer for the careful evaluation of the manuscript and for the constructive and insightful comments provided. The feedback was highly valuable and has substantially improved the clarity, organization, and overall quality of the manuscript.
Below, we provide a detailed point-by-point response to each comment. Reviewer comments are reproduced in black, and our responses are provided below in blue each comment.

Response to Referee Comment 1 (RC1)
This is a nice summary of an innovative educational program that is bringing new skills in data science, machine learning, and cloud computing to hydrologists. Overall, I like the piece and I think it is a nice contribution. I have a few major comments and then several minor comments that I hope will improve the readability of the manuscript.

#We appreciate the reviewer's positive assessment and the detailed, organized feedback.

Major Comments
1. First, I ask the authors to consider revising the sequencing of Section 4. Right now, this is a mixture of materials that the authors have developed and participant data. The ordering was somewhat confusing and could be modified to separate (1) general materials/approach from (2) participant data and outcomes. More specifically, I'm unsure what the point of section 4.3 is – which materials were created by the authors, and which already existed? I had trouble following, and any clarifications would be appreciated. There's also a mix of past/current/future in this section that seems confusing. Could the survey described in Section 4.2 be shared? (Is the information in Section 4.2 necessary to the main text or should it be moved to supporting information?) I defer to the authors on this, but I think there might be some ways to move things around a bit considering the audience they are writing this for to ensure it is as easy to follow as possible. There's a lot of information in this section, and anything the authors can do to organize this is appreciated. Also, consider including a timeline figure that visually displays the information contained in this section and how a typical training "flowed" (and when participants were surveyed, etc.). That said, I defer to the authors on this sequencing and appreciate anything they can do to further clarify the flow of information in this section.

These are the sections, how to respond: Section 4 is the results section of the paper. 4.1 contains fellows selection and the outcome (breakdown of info like us citizen, race, type of university, education level), section 4.2 contains the result of survey done to identify the learning needs, 4.3 contains details of the package developed for the training, it describes the high level detail and presents case studies presented in the training, 4.4 contains the training evaluation like did it work or not based on surveys filled by fellows, 4.5 contains the result from participants post event surveys, 4.6 contains the result of third party surveys on participants, 4.7 contains the result of broader impacts and fellow outcomes.

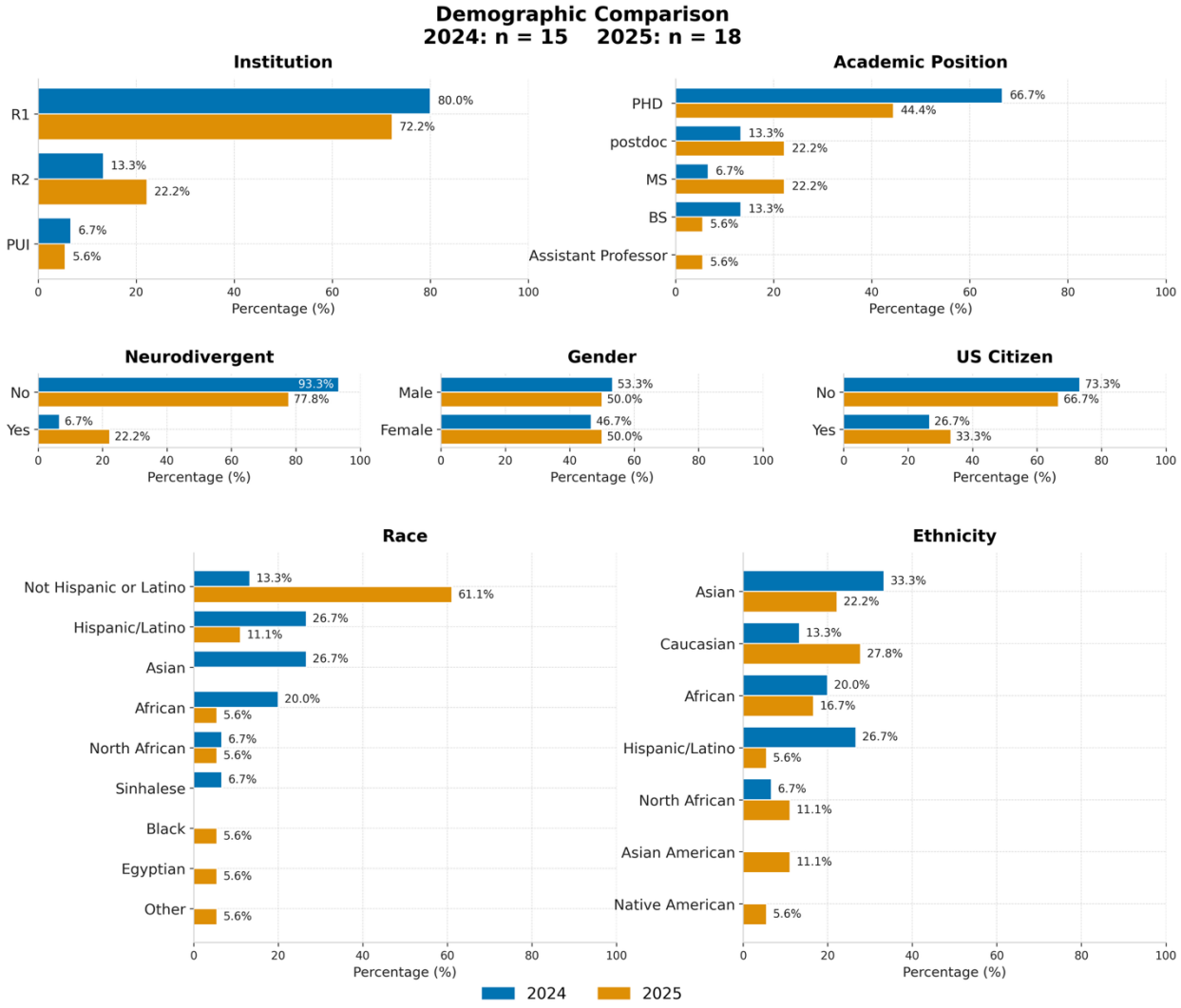
Response: We thank the reviewer for these constructive suggestions in Section 4. We agree that the original ordering interleaved descriptions of the materials we developed with the participant data and outcomes, and we have restructured the section accordingly.

In the revised manuscript, Section 4 is reorganized so that the program we developed is presented first, followed by participant data and outcomes. Specifically, the needs assessment and the training package (former Sections 4.2 and 4.3) appear together as a description of the program and its rationale, while the cohort characteristics and the evaluation results (former Sections 4.1 and 4.4 - 4.7) follow as a single block on participant outcomes. This hopefully removes the back-and-forth between materials and data that made the original sequence difficult to follow.

Regarding the reviewer's specific question about Section 4.3, we have clarified which materials were newly developed for this program and which were adapted from existing resources, indicating the origin of each case study and training component explicitly in the text. We have also standardized the verb tense throughout the section, using the past tense for activities conducted during the study, which should resolve the confusing mix of past, current, and future framing the reviewer noted.

On the placement of Section 4.2, we have considered the reviewer's suggestion but respectfully prefer to retain this material in the main text. The needs-assessment survey directly motivates the design of the training package described in the following subsection, and we feel that separating the two would weaken the logical link between what participants needed and how the program was designed to address it. We have also confirmed that the survey itself is already described in sufficient detail within this subsection, so the relevant information is fully available to the reader in place. For these reasons we believe keeping Section 4.2 in the main text is not problematic and in fact aids readability, though we remain happy to revisit this should the editor feel strongly otherwise.

Finally, we appreciate the suggestion of a timeline figure and have added one (Figure 2). It displays the sequence of program activities: fellow selection, needs assessment, training delivery, and the subsequent evaluation steps and mark the points at which participants and third parties were surveyed. We agree this provides a clearer visual overview of how typical training flowed and complements the reorganized text. We also combined Figures 1 & 2. See below please.



2. The one piece that seemed to be missing, given this is an educational manuscript, were the intended learning objectives for the training. If the authors used learning objectives, please consider explaining/stating these, as much of Section 4 would describe how these are met.

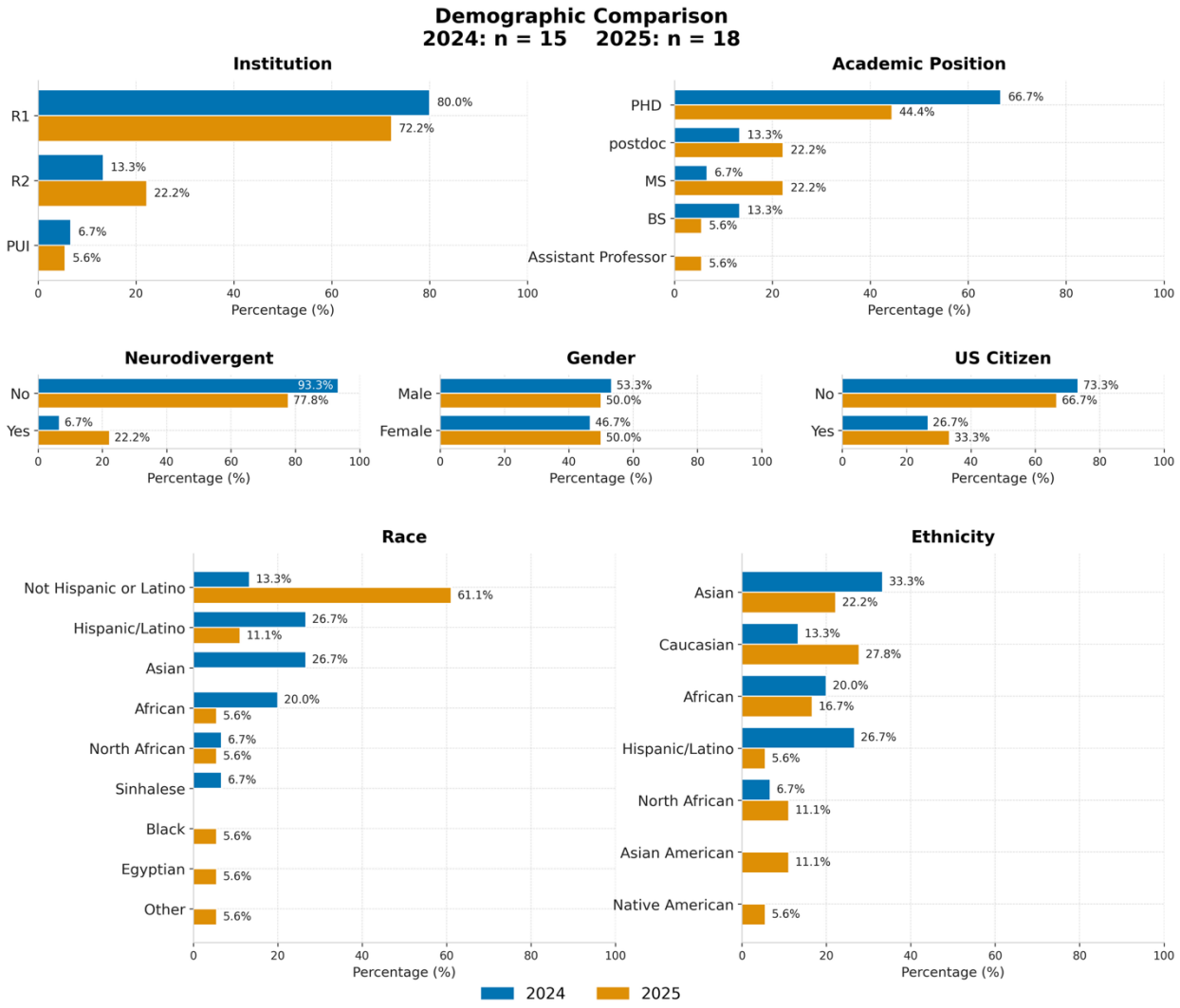
Response: We thank the reviewer for highlighting this important point. We agree that explicit learning objectives strengthen the educational framing of the manuscript.
Accordingly, we have added a new subsection (Section 2.3: Learning Objectives) that clearly outlines the intended learning objectives of WaterSoftHack. See below please

"The WaterSoftHack cybertraining program was designed to equip participants with foundational and applied computational skills necessary for modern water science and engineering research. The learning outcomes focused on integrating hydrological science with data science, machine learning, and cyberinfrastructure practices through hands-on and project-based training. By the end of the program, participants were expected to: (i) develop reproducible computational workflows for

hydrological and environmental datasets using open-source tools and modern programming environments; (ii) acquire, preprocess, visualize, and analyze hydrological, meteorological, and geospatial datasets from publicly available repositories and APIs; (iii) apply statistical and machine learning approaches, including time-series forecasting and neural network models, to solve real-world enviornmental problems; (iv) evaluate model performance using standard hydrological and machine learning metrics and interpret results within hydrological contexts; (v) utilize collaborative cyberinfrastructure platforms such as GitHub, Google Colab, and cloud-based computing environments to support scalable and reproducible research workflows; (vi) understand the principles of cloud and edge computing and their applications in environmental monitoring, real-time analysis, and modeling workflow development; and (vii) strengthen interdisciplinary teamwork, scientific communication, and collaborative problem-solving skills through hackathon and capstone project activities that could be directly transferred to participants' ongoing research, teaching, and professional development."

3. This appears in minor comments, but there are many figures and tables that I think could be easily combined, to centrally locate information for both years in a single figure or table for easier interpretability

Response: We thank the reviewer for this comment. We have merged Figures 1 and 2 to show the fellow details in a single figure. See below please.



4. The Discussion and Conclusions section at the end feels long and somewhat unorganized. Consider renaming Results to Results and Discussion, and including a reflection or lessons learned section here. What are future opportunities in this space, beyond just WaterSoftHack? I'd love to see the authors push beyond recommendations for their own program to some broader recommendations for others, and consider citing some literature in support of this in this section. Furthermore, what materials can others draw from that have been created from this initiative?

Response: We appreciate this valuable suggestion and agree that the original structure blended results, interpretation, and recommendations too extensively.
To address this, the section has been reorganized into a clearer "Results and Discussion" structure, followed by a dedicated subsection titled "Lessons Learned and Recommendations." This new subsection synthesizes transferable insights from the WaterSoftHack experience and expands the discussion beyond the program itself.
In particular, we now discuss broader implications for hydrology and hydroinformatics education, including how modular cybertraining approaches, project-based learning, and collaborative

computational exercises can be integrated into traditional semester-long courses and graduate curricula. We also include additional literature from hydrology education and cybertraining studies to situate our findings within the broader educational research context.
Furthermore, we now more explicitly describe the reusable materials generated through the project, including the WaterSoft Python package, open-source repositories, case studies, computational notebooks, and training modules that can be adapted by instructors and researchers in related educational settings.

Minor comments:
1. Lines 23-24: "The growing complexity and volume of data in water science demand advanced computational skills among researchers, yet significant barriers limit rapid skill acquisition". I would say this isn't just a need among researchers – could this be broadened? Could you be more specific about rapid skill acquisition – for instance, I think you mean skill acquisition to use/analyze/interpret water data. Consider rephrasing this sentence if you can be just a bit more hard-hitting.

Response: We agree with the comment, we revised this portion and made the skills necessity more explicit.

2. Line 24: A transition here would be great "To address this need …"
Response: We agree with your assessment. We have included the transition on the revised manuscript.

3. Line 31: "the first two years" – of what?

Response: Thank you for this comment. The first two years here meant, from two years of conducting the event. We have clarified this in the text.

4. Line 61: The start of this paragraph seems like it is missing out on introducing the concept of point observations before jumping to satellite observations

Response: We thank the reviewer for this comment. We have intentionally kept this a high-level survey of the data landscape rather than an instrument-by-instrument history; a fuller historical treatment of gauging would, in our view, be disproportionate to the paragraph's framing purpose. To make this framing more explicit we have revised the sentence to refer to "USGS-based gauges." No satellite observations are used in this study.

5. Line 61: I'm unsure what 'hydrologically relevant' means here – would just 'hydrological data' work?; 'transformed rapidly' – over what period? Would be good to bring in some of the historical literature on hydrologic observations here.

Response: We thank the reviewer for this comment. We revised the phrase "hydrologically relevant" to "hydrologically related data" to better reflect that several of the datasets discussed (e.g., GRACE and atmospheric reanalysis products) are not strictly hydrological observations, but rather geophysical and meteorological datasets that are highly relevant to hydrologic analysis and

modeling. We also agree that the timeframe of change should be specified and therefore revised the text from "transformed radically" to "transformed over the past century, particularly in recent decades." In addition, we incorporated relevant historical literature on hydrologic observations and monitoring to provide broader context for the evolution of hydrological data systems.

6. Line 76-77: I'm unsure if this is the case. Are there citations to back this up? For instance, I agree that an introductory hydrology course focuses on process theory, but many institutions have more than one hydrology course (though I don't think there's data for this, unfortunately; I would expect this is true at R1s, maybe true at R2s, but not the case at PUIs). Instead, you may consider framing this that if students take an introductory hydrology course, it is focused on these topics. However, many students probably do not take additional hydrology courses and so miss out on this additional training. (I think we have to be careful to expect too much from a single introductory course, which is why I raise this.) I do think there may be a recently published study on hydro informatics courses in the US, which the authors may want to review and reference in this section.

Response: We thank the reviewer for this comment. We have cited recent literatures to back this claim (see Jones et al., 2022; Kelleher et al., 2026). Specifically, Kelleher et al. (2026) presents the survey of undergraduate hydrology courses in North America that supports our statements.

7. Line 96-97: I would say they don't risk being underprepared – it is more likely that they will learn this on the job (or as MS and PhD students, at institutions that either provide training in courses or via research projects). Instead, consider reframing as a positive: Training of this kind will prepare the next generation of hydrologists to work for these agencies and with these large datasets.

Response: We thank the reviewer for their suggestion. As such we have reframed the sentence positively as: Training of this kind can prepare the next generation of hydrologists to work effectively with relevant agencies, leverage large datasets and other available resources, and contribute meaningfully to operational and policy-relevant decision-making.

8. Line 100-105: This section, after the first sentence, feels as though it is the opinions of the authors and so should perhaps appear later in this manuscript (e.g., Section 2)

Response: Thank you for this constructive feedback. We agree that the original text could be interpreted as opinion-based without sufficient grounding in existing literature. In the revised manuscript, we have substantially revised this section by incorporating supporting references from hydrology education, hydroinformatics, and computational water science literature. We revised this portion as

"To address this gap, targeted cybertraining initiatives are increasingly recognized as essential for integrating computational methods into modern water science and engineering education (Wagener et al., 2021). Such programs should extend beyond traditional hydrologic modeling instruction to include the broader computational workflow required for data-intensive hydrological research. This includes (1) acquisition, management, and quality control of large hydrological and environmental datasets (see Jones et al., 2022); (2) data science principles such as statistical learning, visualization,

and reproducible workflows; (3) machine learning and artificial intelligence methods for time-series forecasting, surrogate modeling, and hydrological modeling; and (4) deployment and scaling of models using advanced computing infrastructures (Ruddell and Wagener, 2015; Jones et al, 2022). Recent studies have further emphasized that computational hydrology education should be tightly coupled with real-world hydrological applications and collaborative open-science practices to prepare the next generation of hydrologists for increasing data-rich and interdisciplinary research environments (Kelleher et al., 2026)."

Line 129-130: Again, this feels out of place as though it should appear later in the manuscript (e.g., Section 2)
Response: We thank the reviewer for this comment. We agree that these sentences appear out of place in the introduction section. Hence, we have removed it from the introduction.

9. Line 184: Is this meant to be 'Google Colab'?
Response: Thank you for pointing out the typo. Yes, it meant 'Google Colab'. This is now revised.

10. Line 187: Could you provide a brief introduction to describe what is contained in the next two paragraphs?
Response: Thank you for this comment. We have added a brief introduction to describe the training events here.

11. Line 233: Should this just be civil engineering, water resources engineering? I'm unsure why the dash is included.
Response: We thank the reviewer for this comment. We corrected it as civil engineering with water resources specialization.

12. Line 235: To what field? Data science or hydrology? Or both?
Response: All participant fellows were selected from among the water science community hence it was assumed everyone was familiar with hydrology. So, they were chosen by mixing those with prior experience with data science and those without it.

13. Section 4.2: Is the survey included in Supporting Information? Was the survey pre-populated or open ended?
Response: We thank the reviewer for this comment. The survey was pre-populated with options. We used a few open-ended questions to gather race, ethnicity, and gender data and offer a variety of options in survey responses, including "non-binary" and "prefer not to say."

14. Line 283: This feels like a jump from the previous paragraph. Could you specify how the survey informed the information explained in this paragraph?
Response: We thank the reviewer for this comment and for helping us clarify the transition between these paragraphs. The core training themes of the program including data analysis and visualization, machine learning, and cloud/edge computing were initially identified during the NSF proposal development phase based on prior community needs and discussions within the CUAHSI

cybertraining ecosystem. Following the award, we conducted an additional needs-assessment survey to identify the specific subtopics and technical skills that participants considered most important for their research and educational activities. The survey results directly informed the detailed content of the training modules described here.

15. Line 578-589: Please rephrase this sentence – I think it is grammatically incorrect

Response: Thank you for this comment. We have revised this portion for grammatical correctness.

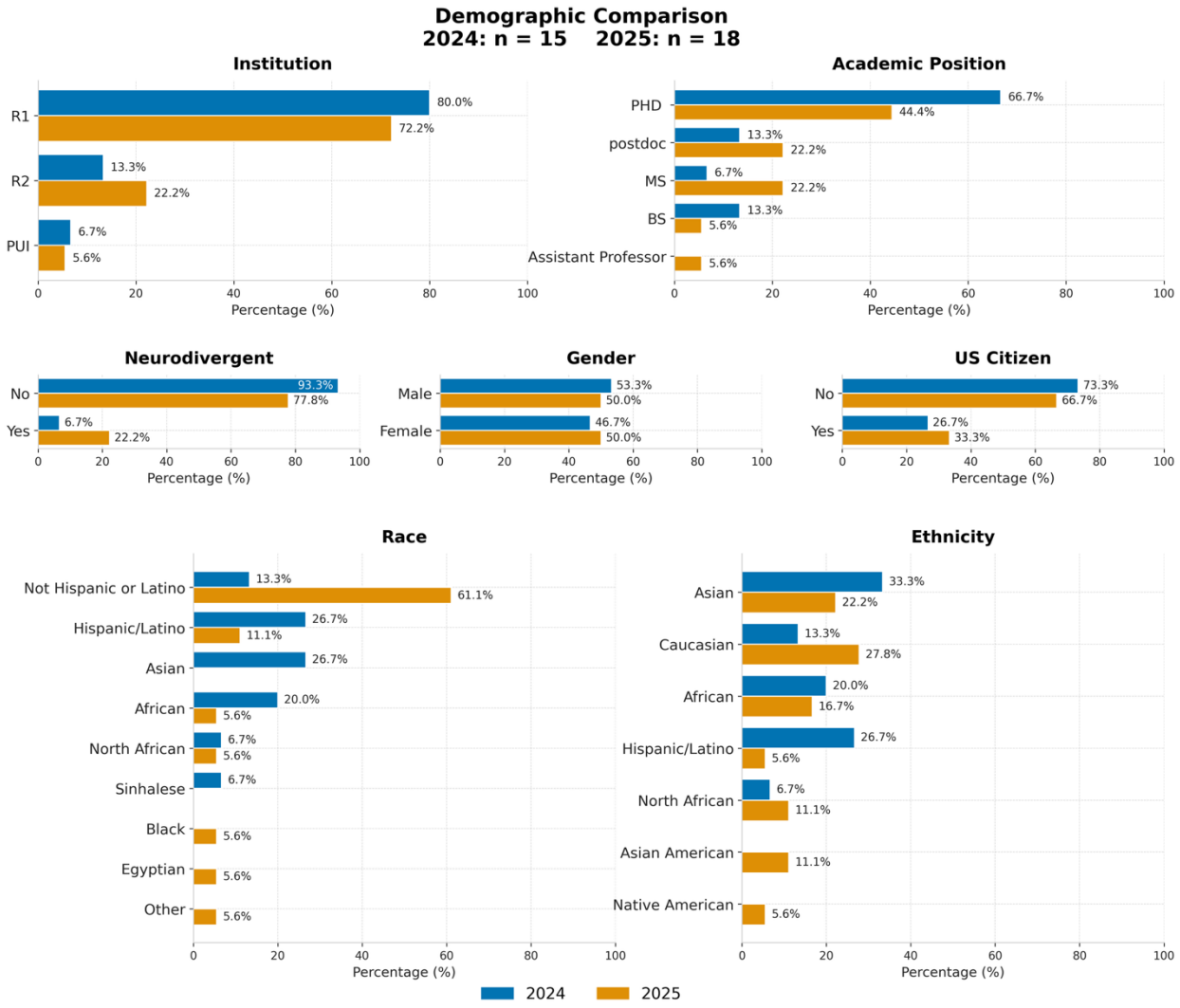
16. Should Section 4 be Results and Discussion? It does feel as though there is some discussion in each of the sub-sections.
I'd encourage the authors to revisit the section titles and revise them to ensure they are representative of what is written in the section – a few seem very specific and only pertain to part of the text within a given section

Response: We thank the reviewers for this comment. We have renamed the current results title Results and Discussion and added an additional section called lessons learned and recommendations. We also revised these portions thoughtfully.

17. Line 688 on: the tense seems off in this section – please revise as needed
Response: We thank the reviewer for this comment. We have revised this portion.

Recommendations for figures and tables:

18. Figure 1 and Figure 2 look quite distorted. Could this be corrected? Consider combining Figures 1 and 2 to show the same observations for 2024 vs 2025 in individual subplots. Color could be used to indicate year.
Response: Thank you for this comment. We have combined Figures 1 and 2 to include participant details in a single figure. See below please.



19. Figure 3 and 4: consider combining into one figure and making an A and B. This will ensure this information is grouped together in the publication

-Consider combining Tables 2 and 3 or into a single figure as A and B

Response: We thank reviewers for this comment. The training events WaterSoftHack 2024 and 2025 were different events, taught different things and are not meant to be compared. Combining them in a single figure forces us to use small font sizes which is difficult for readers. Combining them creates clutter and since they are not meant to be compared, we believe it's good as it is.

20. Figure 10: while this does look cool, I'm not sure it is needed

Response: We thank the reviewer for this comment. We agree that Figure 10 is not essential to the manuscript's main points and have removed it from the revised version.

#We thank the reviewer again for the constructive and insightful comments.

References:

Kelleher, C. A., Gannon, J. P., & Ciruzzi, D. (2026). The current state of undergraduate hydrology courses in North America: A path forward. *Water Resources Research*, *62*(2), e2025WR041736.

Jones, A. S., Horsburgh, J. S., Bastidas Pacheco, C.J., Flint, C. G., and Lane, B.A., 2022. Advancing hydroinformatics and water data science instruction: Community perspectives and online learning resources. Frontiers in Water, 4, p.901393.

Ruddell, B.L. and Wagener, T., 2015. Grand challenges for hydrology education in the 21st century. Journal of Hydrologic Engineering, 20(1), p.A4014001.