Anonymous Referee #1

Author responses are below comments in italics

General Comments

The focus of this manuscript is to explore the effectiveness of working with citizen scientists to collect long-term environmental data with support from a project team, specifically focused on the impact of a small dam removal. In its current form, the manuscript does not clearly align with the stated aims and scope of the journal, it would be more appropriate elsewhere, or requires substantial reworking with major revisions at least.

1. The manuscript is presented as a citizen science focused project, yet there is actually very little detail on the citizen scientists themselves, including how they were recruited ethically, when they engaged with the project, how their engagement was distributed throughout time and perhaps most importantly whether they themselves benefited from the project.

This is a great point. We will significantly expand the introductory section in the revision to address this important piece of missing information.

2. Building on the first point, the manuscript can be read as implicitly attributing data loss or quality issues to citizen scientists, which may be unintended. Careful use of language is required, especially when the authors are claiming that citizen scientists can collect useful data (a message which I agree with).

Thank you for pointing this out, we agree that the implicit messaging is not what we intended. Our aim is to promote the value of citizen science and engagement of citizen scientists and we will revise the language appropriately.

3. Much more detail is needed around the results - the methods section 3.1. is fantastic, lots of great detail but then only d50 and WAV are presented, and it is not clear why. Linked to this, the data analysis and result interpretation requires further detail, for example how were the error bars determined? Did you perform any statistical analysis regarding the difference between d50 at different locations?

Indeed, we did conduct a complete grain size distribution analysis, however, as there were no statistically significant changes, we only reported the d50 results. In the revision, we will add all the other results to the supplemental files. We will also elaborate on the data analysis, including error bars and 95% confidence interval calculations, as this reviewer points out later in the line-by-line comments.

4. The overall structure of the manuscript needs reworking, particularly ensuring that the correct text is placed in the correct location. Introductory text are presented in the discussion (e.g., L329-339), results are shown in L306-316 and L341-350 and so on.

We will do some structure reorganization to the manuscript upon revision, including ensuring that the introductory text in the discussion is addressed. The results the reviewer points out in lines 306-316 and 341-350 are not results, but rather context for the discussion (see comment below for L308-312). In the former, we state the discharge data from the nearest gaging station and for the latter we discuss limitations for data that could not be analyzed. We will elaborate on the latter with respect to the cross section data in the results section when we revise to better contextualize the discussion of limitations.

5. The supplementary data are lacking, with no clear readme file nor information regarding what the data relate to.

We agree that this is a major oversight, particularly the lack of readme file. We have generated a list of files (including a readme) that will be included in the revision, including the excel file suggested by this reviewer below (L230) and additional plots that include the D16 and D84 for the pebble counts.

Line-by-line Comments

L27 "Both species" - what species are you referring to?

Thank you for pointing this out. We will fix this to say: "species of fish and macroinvertebrates".

L36 There are nine dams in the abstract, why are there now seven?

Thank you for catching this. Nine is the correct amount, and we will fix this in the revised manuscript.

L53 "Local streams" - what are the size of these and location relevant to where the citizens are coming from?

For clarity, we will edit this statement to read: "Can citizen science projects effectively monitor wadable streams local to their communities"

L60 By "above-dam levels" do you mean upstream of the dam?

Yes, and we will change this to "upstream of the dam" for clarity.

L66 What about research outside of the US? Such as in central Europe on check dam removal?

Thank you for pointing out this inaccuracy in our word choice. Our meaning was intended to be phrased as "within North America" rather than the language we used that unintentionally omits the rapidly growing body of work on dams and dam removal coming out of Europe. In the revision we will rephrase and include a few sentences to make sure our study is correctly contextualized and that Europe is not omitted.

L78 9177 m is higher than Everest! Assume there is a typo here.

Thank you for catching this typo, the actual mean basin elevation is 1738 m and will be updated in the revision.

L88 Consider adding this information regarding the stream gauge to the Figure 1, alongside the actual data for the monitoring period as this will provide useful context for the rest of the manuscript.

Thank you for this suggestion. Considering this comment and that from the other reviewer, we plan to revise this figure to include the gauge location, upstream dams, and lakes.

L94 Seven dams again - please clarify whether this is correct or not and update the abstract accordingly.

We will correct this as stated above (L36).

L105-114 "This paragraph is a little confusing and needs restructuring. At the moment, it is presented as: "citizen science is good for spatiotemporal scales, yet data quality is potentially not robust. Research has found it to be non robust enough, but this is like comparing apples to oranges, so citizen science data is robust."

I think this needs unpicking more as you can compare citizen science data to that collected as standard, by using the standard data as a reference point. While citizen science has demonstrated value in many contexts, it is important to clearly articulate where it is appropriate and where limitations exist.

Thank you for pointing this out, we agree that the messaging is not what we intended. Our aim is to promote the value of citizen science and engagement of citizen scientists and we will revise the language accordingly.

L125 How were these hours distributed across the study period? Were they equal, or did more monitoring occur in 2018, for example. How did this effect your results? How many people were engaged and how many were the same over the monitoring period?

Thank you for these clarifying questions. The least monitoring hours occurred in 2018 (200) and 2019 (355), and from 2020-2025 the hours ranged from 499 to 896. The number of volunteers follow a similar trend, with an average of 50 volunteers per year. On average, a Stream Team outing tends to have 10-20% new volunteers accompanying a consistent core group. This means that every week there were people who understood the protocol well and new people to teach the protocol to, providing a form of institutional memory for accurate data collection and engaging more experienced volunteers with newer ones. Because of the core group, the differences in hours and volunteers likely did not impact most of the results, especially for pebble and macroinvertebrate counts. We will add these details into the manuscript.

L126-144 This information should all come in further up in my opinion when you are introducing the study site. This would then leave a dedicated section to discuss the citizen science objectives/protocol.

We agree, and we will move these paragraphs to the section on the geographic setting of Rattlesnake Creek in the revision.

L157 This sentence doesn't make sense, please clarify/revise.

We will revise this sentence to read: "Stream Team has monitored Rattlesnake Creek from 2017 to 2024."

L162-163 How were these sites identified? Why only two upstream but eight downstream?

Great question, we will add more detail to the manuscript. 10 sites were selected to ensure weekly visits would cover all sites during the 9-12 weeks of baseline flow during the field season. The sites were chosen so that none are more than a 15 minute walk from a trailhead. Only two sites were located upstream because vehicle access ends approximately 1.5 km upstream of the dam and accessing further upstream reaches would not be feasible within a 3-4 hour long outing. The first site immediately upstream of the dam was chosen to be at a similar elevation and slope to the dam to track any changes with the nearby downstream site, and the upstream-most site is located in the Rattlesnake National Recreation Area & Wilderness, which has stricter access and use rules, making it more pristine. The downstream sites were placed to have somewhat equal distance apart with one site immediately below the dam site and the last one at the confluence with the Clark Fork.

L171 More information is required regarding the width of the streams at the 10 monitoring sites, and why 20-50 intervals were chosen. Did you normalise the stream widths?

Stream widths were taken in 20-50 intervals to balance high resolution and moderate time in the field, as well as to make the math easier to ensure the correct intervals were used. Stream widths were not normalized as we were interested in comparisons between years, not sites. The sentence will be updated to read: "the width of the stream is divided into 20 to 50 intervals to balance resolution and time in the field, as well as allow for simple field calculations"

L230 "Winter season WEN" - this suggests that volunteers had different roles, please expand on this.

We will clarify the different roles by rephrasing: "During the winter season, WEN volunteers entered the data". We will also elaborate on the WEN volunteer recruitment process and roles in the introduction section when we revise.

L230 Can you share the Excel sheet so that others with similar objectives can learn from your design?

Yes, that is an excellent idea. We will add the Excel sheet to the supplemental materials.

L269-271 Why use the upstream sites as comparisons when you have three years of data before dam removal? This is the point of BACI. Furthermore, there might be changes upstream due to flow being no longer restricted.

Thank you for this question. You are correct that the point of BACI is to use data from before and after the event, but it is also about looking at data at impact sites and control sites. We will update the wording of this section to emphasize the importance and role of the data before and after the dam removal.

L273-274 What about the data before/after then as this would give greater detail into how the river responded to the dam removal?

Thank you for pointing this out, forgetting to include this was an oversight. Our analysis includes data before and after the dam removal, and we will add more detail on that to this section.

L275 How many observations? More detail required

We will add the number of observations to the figures. After combining data into groups of above the dam site and below and splitting those into data collected before and after the dam removal, we are left with only 4 observations above the dam before removal, 8 observations above the dam after removal, 15 observations below the dam before removal, and 32 observations below the dam and after removal.

L279 Why did you only calculate d50? If there are 100 measurements per sample site why not include the entire grain size distribution, including d16, d50 and d84 as stated in the methods?

We did calculate d16, d50 and d84, however, because there was no discernible trend, we chose to only show d50 in the figure. We have a couple of ideas to address showing all of our data such as including all the figures in the supplemental material or creating a new figure using box plots to more efficiently display the results of the grain size distribution analysis.

L285 Figure 2 - I think that these figures would be better combined, showing the change throughout time based on the year. It is unclear where the error bars have originated from, why not use a percentile metric to calculate the 95% confidence intervals? See Eaton et al., 2019, ESurf, https://doi.org/10.5194/esurf-7-789-2019 for example.

We are not entirely sure what the reviewer is asking for here. In Figure 2, graph a shows the d50 distribution per site per year and graph b shows the d50 distribution per year for upstream and downstream sites (all combined). If we make the box plot (see comment above) that shows the grain size distribution, this would replace graph b in Figure 2.

If we don't use the boxplot for graph b, we will explain how the error bars were determined in the results and revise the caption. The error bars represent one standard deviation, but we will update this to either use a percentile metric, standard error, or a 95% confidence interval.

L289 Figure 3 - similar comment to Figure 2, I don't think separating the data in this way is additive, therefore please think about combining the plots. Also it's unclear how the error bars were calculated.

As with the comment above, we are unsure if the reviewer is suggesting to remove graph a (showing individual sites) from the figure, leaving only graph b (that shows the upstream and downstream sites). We think that if we make the box plot for grain size distribution as described above, we would change graph b in this figure to a box plot as well for consistent visual representation of the data. Similarly, if we do not revise to include a box plot, we will be sure to include the error bar calculation method.

L293-299 This paragraph would benefit from reframing, as it reads as if the citizen science data were effective in their collection, however a third of the data collected was not usable. Please consider how this is framed.

We agree, the framing of this paragraph unintentionally puts the citizen scientists at fault when it is the managers' responsibility to ensure that datasheets were complete before leaving the site. We will work on wordsmithing this section to emphasize the importance of field data checking to ensure that protocol has been executed and reduce the emphasis on missing data not collected by the citizen scientists themselves.

L301-305 This is not discussion as it doesn't link to the presented results as there was no significant change.

This is a great organizational point and we completely agree. This paragraph belongs in the results section.

L307 "sediment flow" - you have not reported measuring sediment flow, rather grain size distributions and WAV. Please reword.

We agree that this can, and should, be rephrased, which we will do in the revision. For example, the word "flow" can be replaced with the word "distribution".

L308-312 This is results rather than discussion, why isn't this presented earlier in the results section? It contains important information that has not been presented.

Since the citizen scientists did not measure discharge (these data come from the MDNRC gaging station), we did not include the discharge values in the results. These values are not site specific (we were unable to get site discharge without the cross-section data), we just used this downstream gauge for context of the variability of discharge of the creek before and after dam removal.

L312 What irregular variation? More information is required.

Good point, we can add a sentence to explain the variation in the sampling (and its irregularities) to clarify. This is in reference to the irregularities in data collection arising from individual differences of the different groups of people collecting the data each time.

L319-324 This is introduction to the area rather than discussion. Equally for the sluice gate statement, this should come earlier in the text.

This is a good point, we should definitely introduce the recreation usage of the area early on and will do so in the revision. The sluice gates are already introduced in section 2.2 and discussed again in section 3.2 before they are mentioned again in the discussion section.

L329-334 This is not discussion, it is introduction. Secondly, the results presented don't back up these claims as there's no evidence to support them. It's also not what the focus of the manuscript currently is. If you have anecdotal stories/evidence that would be more effective.

We are planning on increasing the introductory material on the citizen scientists for the revision. This paragraph is entirely anecdotal, communicated to us through personal communication from the WEN director. We are going to have further discussion with WEN and will solicit more specific anecdotal

evidence at that time such as direct quotes from participants. We hope that participant quotes are what this reviewer means regarding supporting evidence.

L336-339 Again this is introduction, not discussion.

This is material that we intend to include in the revised introduction section about the citizen scientists, recruitment, program, etc. That revision should address this concern.

L341-350 This raises a quality control issue that has not been addressed - why was all of this potentially fantastic data not quality controlled or checked regularly?

That is a great question. As independent scientists working with data from the citizen science program at WEN, we had no control with respect to data quality control. Our conversations with WEN suggest that cross sections may not have been as high of a priority on Stream Team outings for early analysis, which may explain why this error on data collection sheets was not caught during earlier years. Earlier versions of the data entry sheets in the field did not have clear spots to record all the data (i.e. distance along the cross section). This is also one of our take home messages from this paper: citizen scientists do a great job on what they are instructed to do, but if they are not instructed to do something and/or if there are not places in datasheets to record data, citizen scientists likely will not know to collect those data.