Authors' Response to Comments

November, 2024

Below are the authors' response to a second round of reviewer comments. The original comments were reviewed by the authors and a response was submitted to the journal on August 2024. There are some new comments that continue to improve the manuscript but there is also some confusion as to the goal of this study which leads to reviewer comments that are less beneficial. The purpose of this study is to assess the precision and accuracy of three commonly used sediment modeling techniques for post-wildfire mitigation. The study used empirical field observations to make this assessment possible. The results are useful for evaluating the utility of each of the models as well as for determining areas for improvement (e.g. hillslope gully process is not evaluated by WEPP, MUSLE was not originally intended for non-equilibrium post-wildfire conditions, and WARSSS relies on a proprietary set of models that provides similar results as WEPP at the watershed scale). This study adds to the state of the science for post-wildfire sediment forecasting as well as determining areas for sediment mitigation projects. A relatively novel sediment mitigation design (an alluvial fan restoration) is provided as an example of how these sediment forecasts can be successfully used to reduce impacts to both human and ecological values. The purpose of this study is of interest to an international audience and is similar to many other papers that validate (or invalidate) a particular model or theory for forecasting or predicting an environmental phenomenon.

This submission includes a track changed version of the manuscript incorporating all reviewer comments as well as a "clean" version of the manuscript.

Reviewer One, second round of comments (initial authors' response in August 2024)

Authors' response in bold type the reviewer comment is in regular type, this response build on the previous round of reviews.

Comments on Revision

The authors have addressed some of the comments from the last revision, but several major concerns remain with the existing manuscript. First, some of the literature citations are inappropriate. For example, they discuss sedimentation modeling and use a reference to a hydrology-only paper that does not address sedimentation directly (see line comments below). Second, the authors failed to address the comment about a timeline of events, and this remains important. This could be done relatively easily with a few additions to Figure 9. For example, in addition to showing the sediment, the authors could simply add a secondary axis with rainfall intensity that shows the timing of the major storms. Crucially, I think the authors need to also show the timing of the completion of their "alluvial fan restoration areas" on Figure 9 as well. The third and final concern, is that the authors do not take the opportunity to explicitly compare observations with modeling. For example, everything in Table 1 appears to be a modeling result, but they have an opportunity to compare the model with observations. Same comment with Figure 9. That shows sediment observations, but it'd make sense to compare these with the modeling. I realize that the modeling and observations are obtained over diKerent timescales, but you can sum the observations to the annual time scale and compare that to the modeling. Same comment with Table 3. Why not add a column that says "Sediment Retention Observed at Channel Design".

The literature citations have been reviewed and revised accordingly (see line comments). The timeline of events is addressed throughout the paper, Figure 9 now shows flood events with the empirical sediment observations, Table 1 now includes the sediment transport results as well as the observed sediment, Section 2.1 now includes a brief discussion of the overall timeline of flooding, empirical observations, modeling, and mitigation structure development. While this improves the case study aspect of this paper it paradoxically is opposed to both reviewers' comments about improving this paper to provide a more clear scientific benefit. The purpose of this paper has been, and continues to be, to evaluate the efficacy of different sediment models for forecasting sediment risk and designing sediment mitigation strategies (as stated in the abstract and at the end of the introduction section). We believe the paper will help both applied flood management as well as the state of the science by demonstrating the utility and precision of various modeling techniques. Hopefully this will also help with refining these models (especially the easy-to-use WEPPcloud) to be more precise and more used internationally. Focusing on the case study timeline, or other details of the case study, distracts from the purpose of the paper. The purpose of presenting the mechanics of the case study is to provide enough methods to allow for the tested hypotheses to be replicated if so desired, other details beyond this objective are tangential to the paper.

We strongly disagree with the reviewer's third concern that the authors do not explicitly compare observations with modeling. Here are some example areas where the comparisons are made:

From the abstract: "*Empirical evidence from four floods in 2021 indicated 9,900 Mg of sediment yield to city of Flagstaff neighborhoods, WEPPcloud estimated 3870 Mg/year, MUSLE predicted 4860 Mg/year (based on the four events), and the WARSSS suite of models predicted 4630 Mg/year. Both WEPP and WARSSS estimated more sediment yield from channels than hillslope (51%/49% and 60%/40% respectively) though the spatial patterns differ between the models.***"**

From the Results: "*The commonly used WEPP model demonstrated much lower sediment yields (3870 Mg/year) than the WARSSS model (4630 Mg/year) and empirical results (9900 Mg/year in 2021) for the Museum Fire burn scar and Spruce Wash watershed and slightly less than the event based MUSLE model (4860 Mg/year).***"**

From the Discussion: "*All three modeling domains, MUSLE, WEPP, and WARSSS showed drastic increases in channel and hillslope sediment yields post-wildfire in this case study. Both WEPP and WARSSS predict slightly more sediment yield from existing channels than from hillslope processes, however the hillslope gullying and rill erosion is substantial. The similarity between model results, and less than an order of magnitude comparison with empirical results, indicate that both WEPPcloud and WARSSS are useful for sediment predictions.***"**

From the Conclusions: "*This case study shows the utility of both WEPPcloud and WARSSS for predicting sediment transport to the city of Flagstaff, Arizona. The agreement between both models for sediment transport, and within an order of magnitude comparison to empirical observations from flood events in 2021, is encouraging. The difference between models was largely in the spatial pattern of sediment yield. Both models indicated a slightly higher contribution from channels than hillslopes but WARSSS, because it is partly empirically based, was better at identifying "hot spots" of both channel and hillslope sediment yield.***"**

The results section provides the quantitative comparisons, the discussion section provides a commentary on the results, and the abstract and conclusions both summarize those comparisons (as is the format with nearly every peer-reviewed scientific paper). The comparison of the field observations with the modeled results is the focus of this paper, it is with some dismay that we read that the reviewer does not agree that enough to time is spent in this paper providing explicit comparisons of the modeling results to the empirical results. We have attempted to make the Introduction and Discussion even more explicit in the comparisons in this round of review but the paper is already written fairly straightforward.

Right now my biggest concern is that this paper is a site-specific case study that does not have large application outside of the specific scenario. Moreover, I'm concerned that because the modeling is not compared to observations directly it's hard to evaluate how well these models actually estimated what was observed. Finally, I do not understand how the mitigation structures influenced/did not influence the sediment retention because I do not know when they were built with respect to the storms.

As shown above there are ample comparison between the modeled results and the field observations. This paper uses a case study to demonstrate the utility and precision of three modeling techniques, that is the purpose of the paper (as explicitly stated in the Introduction) and has international applications as all three models can be used globally. Additionally, the WEPPcloud interface is free and extremely user friendly, any study that provides precision and accuracy results for this model could be very helpful to the larger flood risk community due to the lower barrier to entry on this model. The paper has been further improved (see line comments) to make the utility of the results even more explicit. Finally, in regards to the mitigation structures, this is a secondary purpose of the paper and a minor portion of the paper. The purpose of including the mitigation structures in the paper is to show the utility of the modeling techniques in designing and selecting locations for mitigation structures. We provided more detail on the "alluvial fan restoration" strategy as an introduction to the technique, since this case study used a relatively novel mitigation strategy. Line comments requested more references to this

technique (mostly in regard to the Rebecca Beers conference paper reference) which we have attempted to improve upon. Since this technique is relatively novel there are no peer-reviewed articles that go into depth on the alluvial fan restoration technique and we are forced to use conference abstracts and technical reports. Again, the exact mitigation structure design is a minor component of this paper.

Line comments

29 I understand what you mean by sediment forecast here, but I don't understand what you mean by sediment forecast on line 28, where you say that "Sediment mitigation structures … are discussed as real-world applications of sediment forecasting…". I think this might just be a problem of language imprecision. I think you mean something closer to: "We discuss the real-world implications of using models to make sediment forecasts and using modeling results for the design of sediment mitigation structures." Either way, please refine this sentence because right now it sounds like you are saying that sediment mitigation structures are provide sediment forecasts, which doesn't make much sense.

This sentence has been revised to provide clarity.

34 Provide references to support this assertion, especially references that indicate that it is an "increasingly important issue", or change the language if that isn't something that is supported by research.

The rise of catastrophic wildfire, especially at the WUI is very well documented. To prevent adding to the already long reference list we re-used existing references that document the rise in impacts as examples (specifically Ebel et al. 2023 which addresses it in the first sentence of their abstract and Sankey et al. 2017 which address it in the first two paragraphs of their introduction section).

36. Sankey et al., 2017 is not a paper about flooding, and therefore you should adjust the reference. The word flood is not even used in that paper.

The incorrect reference was used here, updated to Sankey et al. 2024 and an additional reference (Kinoshita et al. 2016) is added.

40. Ebel et al., 2023 (which doesn't have a year in your references) is about hydrology and does not specifically address these concerns: "damaging debris flows and sediment sourcing, transport, and aggradation". Here and elsewhere, please use references that support the assertions you are making.

The Ebel et al. 2023 paper discusses sediment sourcing modeling by calling it "erosion". Multiple models are discussed including WEPP, but we understand that one would have to read this paper comprehensively to parse out the sediment models from the hydrologic models. Ebel has been removed at this line location and two other more

clearly defined sediment model references have been included (Moody et al. 2013 and Smith et al. 2011).

46. Same problem as the last two. These references aren't really well suited for the points you are trying to support.

These three references were provided as examples of the large body of knowledge on sediment impacts on watershed ecosystem recovery.

51. Either I'm confused about what you are referencing with Ebel et al., 2023, or you are confused. But I think you are trying to refer to this paper that is specifically focused on hydrological modeling and does not touch on sediment modeling at all. Is there a chance you are actually referring to a diKerent paper?:

Ebel, B. A., Shephard, Z. M., Walvoord, M. A., Murphy, S. F., Partridge, T. F., & Perkins, K. S. (2023). Modeling Post-Wildfire Hydrologic Response: Review and Future Directions for Applications of Physically Based Distributed Simulation. *Earth's Future*, *11*(2), e2022EF003038.

The three instances of the Ebel reference in this paragraph have been removed, The Ebel paper does discuss sediment source modeling through their "erosion" discussion but we understand that this may be nuanced and have removed the reference to prevent confusion.

62: Here you reference a conference abstract by Beers et al. 2023, and there is no mention of a loss in stream power or accretion upstream of neighborhoods in that conference abstract. Also, you inadvertently changed the title of that conference abstract, so I suggest you change it back to the original title. In any case, I don't think that the information in that abstract can be used to support the assertion here.

We removed the Beers reference and added two other references (Grover 2021; Rosgen and Rosgen 2015). Since this mitigation technique is relatively novel we are forced to use conference proceedings and technical papers for any prior reference to the technique.

103 When you say that the flood events allowed for empirical comparisons to the modeled predictions be more specific. Do you mean you compared sediment discharge, volume, mass, flood velocity, etc? Specify to readers what exactly could be compared.

Added verbiage to explain that we are comparing sediment discharge.

119: a 10-100 time increase in surface water runoK compared to what? Mean annual flow? Also, by runoK, do you mean discharge? Or depth? Be specific here.

The wording has been changed to "peak discharge", there is no mean annual flow in these ephemeral channels as many years there was no flow at all. This only changed post-

wildfire, and even then flow duration is still measured in minutes and hours.

126 There is something I don't understand about this: "Areas downstream from high sediment yield areas were identified as "work areas" …". I thought that the "work areas were specific mititation areas that were defined. This makes it sound like any area downstream from any other area with a high sediment yield is a "work area". I think you need to clarify this, and locate these areas on a map.

This sentence provides the method of determining "work area" site selection. We added slightly more clarification (low gradient areas…). These areas are shown in Figure 6 in the Results section, a more appropriate section as the Methods should not be showing results.

148: International readers of this European journal are likely not concerned about "who" estimated discharge, but I think they will be concerned about "how" it was estimated. So please include that information.

The sentence has been revised to remove the name of the "who" and provide a little more context to the method. The full 2-D modeling method is quite long and complex. The references within the sentence at this line number provide the needed detail for recreating the discharge estimates.

152: Are you saying that the floods incised the channels and then they will widen over time. Or are you saying that they were already incised prior to the fire, and will widen due to flooding. Please clarify.

While the mechanics are not important (either of the interpretations of the sentence will provide the divergence in sediment models explained above this line item) we have clarified the sentence to explain that the channel incises post-wildfire and then widens into the future.

154: here and elsewhere, when you are describing methods you have used, please use past-tense. For example, you say: "Sediment transport estimates are …", but that activity is over, so you should say "Sediment transport estimates were …". Same comment on 165 "sediment transport analyses are…" should be "sediment transport analyses were…"

We corrected the tense in Line 154 but not at Line 165. At Line 165 the statement that sediment and flow data is difficult to collect for ungauged ephemeral streams is still relevant now and into the future. It is not a past-tense one time statement.

207: Here you need to say the version number of the software, and provide a reference in your works cited.

Added.

214 say why you chose 1 inch, 2 inch and 3 inches per hour. Also, use "in" rather than a quote to abbreviate inches, as this will be more clear for an international audience.

Rewritten to be more inclusive (better convey "inches"), the rationale for using standardized design storms is rather apparent and further explanation is not needed to explain this study. The manuscript is already fairly dense on the details of the methodology. Methods is meant to provide the reader the ability to recreate the study and to understand any weakness in the study design. Explaining the rationale for using every detail of a model is beyond what is needed to provide the study design to the reader.

223 I don't think you can use these words together " measured qualitatively" because if it's qualitative it's not actually a measurement. So change to something like "estimated qualitatively". Also, specify the diKerence between sediment and debris here. Is debris wood and trash? Say how it's diKerent than sediment, or just remove the term debris.

The sentence has been refined.

234 after "…average annual sediment transport…" add units (e.g., Mg/yr) so readers know the units you are using.

Added.

243-244 I don't think you can justify this choice. In alluvial rivers the two-year return interval is often substituted for Bankfull, not the one year RI. Also what do you mean by "post-wildfire channel forming discharge" I don't think that's a concept that's been shown. What would be diKerent about a postfire channel forming discharge versus a non-postfire channel forming discharge? Why is "channel forming" relevant here? If you think these choices have merit, then I think you need to explain them or point to literature that explains them.

This comment doesn't make sense, this paragraph is outlining how the model (in this case FLOWSED/POWERSED) was set up for this study. We don't have to explain the rationale for the individual model inputs, just what the inputs are in case a reader wants to re-create the experiment. Model parameters will be different for other watersheds around the world. Also, this case study is not on an alluvial river but on ephemeral mountain drainages that are rapidly adjusting to a watershed regime change in terms of soil, vegetation, and channel morphology character, a "typical" 2-year recurrence period for bankfull is not appropriate in these highly unstable high gradient systems.

257: You need to provide this information in a table somewhere and then reference it here: "bankfull cross- sectional area, Manning's n value, bankfull discharge, slope, suspended sediment (mg/L), measured bankfull bedload (lb/s), a flow duration curve, and a sediment rating curve comparison"

Since this data includes rating curves, sediment assemblages (ranges), and discrete site data it would be difficult to represent in a simplified summary table. A reference to the technical report is now provided instead. The Appendix (B) of this technical report includes 79 pages of supporting data that "feeds" the FLOWSED/POWERSED sediment

model. Similar to our comment above these model input values are only useful for recreating this case study and are not integral to understanding the precision and accuracy of the various models being tested in this study.

263 I'm a little lost on how you would use a "bankfull flow" if you are trying to estimate flow on an alluvial fan. Please explain.

This is explained in these two paragraphs in the manuscript but a little extra clarification has been added. The bankfull flow is used in the existing condition to calculate a baseline sediment load for the "work area" project. The "work area" (or alluvial fan restoration) is then modeled as a proposed condition to determine the benefit, in terms of reduced sediment flux. Bankfull flow is not determined for the proposed condition, the same flow rate is used as the existing condition (which is determined by bankfull calculations).

285: Insteady of saying "G" channel, which is not a universally understood metric. I suggest you use a few short descriptions of the channel type. Something like "the channels are riKle-pool channels defined as "G" in XYZ scheme".

Some clarification has been added for both "G" and "D" classifications. Both are "classic" incised (G type) or braided sediment laden low gradient (D type) channels.

306 Avoid contractions like "don't" in scientific writing

Corrected

367 Earlier you said there were four events. Please correct or explain the discrepancy. On line 24 it says "MUSLE predicted 4860 Mg/year (based on the four events)"

It was four events, this sentence has been corrected.

374 Again the reasoning for these choices of rain events were not justified. Please state why you chose those particular storm amounts/durations.

These were the actual storm events in 2021.

398 State if field observations confirm the modeling here. I think that readers will want to know if the modeling matches what was observed at those cross sections.

The alluvial fan restorations/"work areas" were not established until after the 2021 flow events. There are two timelines here: the modeling and verification through field observations, and the modeling to inform mitigation measures. This will be clarified in the Figure 9 timeline as recommended at the beginning of this review as well is in Section 2.1.

408 I think the critical question here is how do the modeling results compare to field

observations. You have an opportunity here to compare model predictions with field observations, and that's the critical piece that seems to be missing.

The modeling results compared to the field observations is provided multiple times (Abstract, Results, Discussion, and Conclusion). Unfortunately, we do not have field observations for the sediment transport/mitigation structure side of this study as the mitigation structures were built after the 2021 events. This is explained in this section ("Flow events in 2022 were muted in Spruce Wash due to small rain events, the alluvial fan sites that were constructed prior to monsoon season did appear to function well in terms of sediment aggradation and attenuation (Figure 10) however there were no flow events that over-topped the channel within the city to provide empirical comparisons.")

This is now furthered clarified in the Figure 9 which will provide a clear timeline of flood events and also in Section 2.1 which describes the flood events and the modeling and mitigation construction activities.

416 This mention of the Pipeline needs more detail. How was the 70-80% measured? Was it repeat lidar, photogrammetry? Also, I think a critical piece is the timing after fire. Retaining 70-80% of sediment several years after the fire doesn't really say much about how the sediment retention structures would behave immediately after the fire. It is well known that wildfire sedimentation is typically highest in year 1 and goes down drastically in the following years as the watershed recovers.

Since the authors did not conduct the field observations we cannot comment other than it was "repeat surveys" and sediment haul off (as referenced in the paper from Tiffany Construction LLC and the Coconino County Flood Control District). The Pipeline Fire impacted nine (9) watersheds so it is likely that multiple methods were employed in their estimates. We added timing considerations, however we must caveat that substantial sediment is observed after fires in this part of the world. Regular flooding and high sediment loads are still observed from other fires in the area of this case study. This is largely due to the vegetation state change in many of these areas (change from ponderosa pine to grassland). An in-depth discussion of post-wildfire watershed state change is outside the scope of this paper, but it is an interesting topic.

450 What "hydrologic forecasts" are you referring to here?

These are presented in the next two sentences with four references to the papers/reports that provided their predictions for future water yield in the project area.

501 I don't think that prior research supports this assertion that there will be "…substantial sediment loading for the foreseeable future…" it is well known that sedimentation rates after fire decline precipitously after the first year after wildfire. Within 3 years it is very unlikely that you'll continue to have fire-related sedimentation problems. Provide evidence that would support this point if you want to make this assertion.

The immediate post-fire sedimentation may decrease however these channels are now incised, the hillslopes now contain substantial gullies and rills, and the vegetation has transitioned from dense pine forest to sparse grassland. Studies in the region show

elevated sediment and flood risk for decades after a fire not to mention geomorphic adjustments. Geomorphic references are provided and we provided two new references that indicate uncertainty in long term sediment loads (McGuire review) and a comparison of flow events (which can function as a surrogate for sediment transport, from the 1977 Radio Fire and this current fire: JE Fuller 2024 technical report).

507 Here and elsewhere, suggest replacing “poor conditions” with something else (e.g., transient or non-equilibrium conditions). I think you are trying to say that there is erosion or change, but calling it “poor” implies a judgement call on what is good/bad, which is really dependent on the observer.

Replaced three instances of “poor” with “non-equilibrium”.

514 Again provide evidence of how you measured 70% sediment retention on the pipeline fire. Also, I just looked up the Pipeline fire and I can see that it burned starting in June 2022. Did you have alluvial fans built on that fire to capture sediment in 2022? If so state the date of when those sediment retention structures were built with respect to the date of the fire and storms.

As mentioned in the manuscript this was not studied or observed by the authors, this was an observation of our County (regional government) counterparts at the nearby Pipeline Fire. There is a large body of applied research on the larger Pipeline Fire, including the efficacy of the alluvial fans, but to our knowledge none of that data has been published in a peer reviewed outlet at this time.

526 Again, the Beers et al., 2023 reference is just a conference abstract so there is no additional detail to be found on this.

See comment immediately above, the presentation is available upon request from the Arizona Geological Survey (Rebecca Beers). To clarify the data source we have added “personal comms. Rebecca Beers”. Unfortunately, none of this data is currently published in a peer-reviewed journal (this current manuscript is for a fire three years previous than the Pipeline Fire and we are still in peer review). As an aside, one of our motivations for publishing this work is to provide techniques and data to the scientific community for both practical application as well as building the body of knowledge on post-wildfire sediment mitigation. Traditionally much of this work is completed by engineers and applied scientists who have little financial or professional incentive to publish their work (and in some cases publishing can be detrimental to their careers if the results are not to the liking of the funding agencies). There are probably a dozen papers worth of data, lessons learned, and techniques from recent fires in the Arizona, USA area alone, little of it will be preserved through publication unfortunately.

527 Quantify what you mean by “moderate” and “long-term average”? Do you mean you retain 50% of the sediment or 98% and is this 50% of the long-term average measured using some dating technique or something else?

We have revised this statement to demonstrate that these were field observations and not quantified. The authors have observed these structures working in multiple post-wildfire environments including the 2010 Schultz Fire, this fire (2019 Museum Fire), and the 2022 Pipeline Fire. Averaging the efficacy of the structures over multiple years we qualitatively observe that the structures retain a considerable portion of the upstream sediment load but that much fine sediment still passes through. Further investigation is obviously warranted but is not the aim or scope of this study (we are focusing on the model efficacy in this paper and how the models can inform mitigation design). The purpose of this sentence is to explain that we have observed these structures working and then transition into the next sentence caveating that high gradient alluvial fans might require maintenance during high intensity floods.

Figure 4. Somewhere, maybe in a supplement. Provide definitions for all of these channel types.

The caption already references the "Rosgen classification system" which is readily available, for free, on the internet through numerous venues. A general description of the channel type is also included in the caption to provide the reader with enough context to understand the watershed level pattern. Adding a supplement to this paper would be unnecessary.

Figure 9. Please add rainfall intensity as a secondary axis on this so readers can compare the sediment mass per day with the rainfall intensity. You mention the four storms, but it'd be helpful to see where those storms exist in time compared to the sediment yield. Please also indicate the date that the alluvial fan restoration areas were completely constructed. You can do this with a single vertical line on the completion date.

We reworked Figure 9 to show the peak flood flow at the downstream end of the modeling domain and the beginning of the built environment (the city neighborhoods). This provides the best comparison with the empirical observations of sediment. Rainfall is provided elsewhere in the manuscript and has been simplified into an average over the burned watershed (there are four rain gauges within the burned footprint). Adding rainfall as a third axis would make the figure overly "busy" while not providing any new information.

The alluvial fan "work areas", or restoration structure, were designed and built based on the models and were not built until 2022 and early 2023. The timeline will be explained better in the introduction or methods to show the difference between the sediment forecasting (conducted in 2021 and 2022), the empirical observations (conducted in 2021), and the construction of the mitigation structures (conducted in 2022). Please see Section 2.1 in the revised paper.

Reviewer 2, response to second round of comments (initial response on August 2024)

Authors' response in bold typeface.

The authors have addressed some of the comments raised by the other reviewer and myself. Although this new version is clearly an improved piece of work, I think that this manuscript still needs some work before it can be accepted.

Overall, the context of this research is still too narrow. I would recommend the authors to zoom out from their (nice) case study and think of a broader audience. My comments on the abstracts is an example of such a broadening. The key point in an abstract is not necessary to provide details on model's name, but instead to say that several physically-based models have been run and compared to asses... Idem for the sediment yield values. What interest a reader is also to know if the values are normal, high; expected, similar to other cases? (etc.) Without such a broader perspective, the international dimenion of this research is missed in some way.

The reviewer is missing the purpose of this study. As mentioned in the August response this study is intended to assess the precision and accuracy of commonly used sediment models using field observations from actual flood events to provide a measure of both precision and accuracy. The purpose is not to provide a measure of sediment load to compare to other post-wildfire environments. As mentioned previously the sediment load will vary based on numerous environmental factors (watershed size, burn severity, climate, soil, etc) that would make this study rather inconsequential if our purpose was only adding one more data point at a global level. Paradoxically the reviewer is requesting that this study become "merely" a case study by focusing on comparing sediment load with other case studies instead of comparing modeling techniques using actual flood events (a much more interesting study and more applicable globally). There are plenty of examples of published papers that validate, or invalidate, model predictions based on field measurements, this paper is written in the spirit of that line of scientific inquiry.

The authors of this paper have collectively worked dozens of post-wildfire assessments in the western USA and frequently have policy makers, land managers, scientists, and hydrologists ask if a certain sediment model is providing "correct" results or if the post-wildfire mitigation strategy (or strategies) are based on accurate modeling. The value in this study is showing that both WARSSS and WEPP are in general agreement between each other and within an order of magnitude of observed events. The WEPP model, in particular, is of interest to many as the WEPPcloud interface is extremely easy to use and has seen a great increase in use globally. Having a study that describes the precision and accuracy of this low-barrier-to-entry modeling regime is helpful not only for researchers in developed nations but also for applied science and management in areas with limited financial and engineering resources.

Overall, I believe that many of my first round comments (and also those of the other reviewer) remain valid.

Smaller comments to clarify some of my requests/comments:
The study area maps would benefit form an information about the topography. I mentioned adding elevation quotes on the maps to have a broad ideas of the differences of elevation between the

summits and the valley floors.

This was addressed in the August 2024 review, elevation is provided in the body of the text for some key locations. Adding elevations to the maps would make the maps difficult to read and also would provide nothing useful. The scale of the maps is large enough that any elevation label would necessarily span hundreds of meters of both vertical relief and horizontal distance, making the label useless. Modelers who will be reading this paper will be using similar digital elevational datasets as are referenced in the Methods section. The repeatability of this study hinges on those public datasets and not on elevational labels on relatively small maps within figures. Latitude and longitude is noted on each of the maps for quick reference in whatever digital elevation model (DEM) that a reviewer or reader would like to use for evaluating this case study. Paradoxically this repeated request for elevation data is contrary to the request that this study become less of a case study and more applicable globally.

August 2024 comments to a similar request by the reviewer:

Unsure of where the reviewer would like elevations called out, the topography is highly heterogenous. Example elevations for Mount Elden, Dry Lake Hills, Mount Elden Estates neighborhood, and Paradise/Sunnyside neighborhoods are now included in the Study Site section. Contour lines, and DEMs, for the area are freely available online at multiple sources (e.g. USGS National Map, Google Earth, City and County GIS portals, etc).

In the discussion, gully erosion is mentioned. My comment on that was that a process that is discussed but not even mentioned on an earlier stage is something that is awkward.

This was addressed in the August 2024 review, gully erosion is a component of hillslope erosion, a thesis on gully erosion is not the topic of this paper. Verbatim from the August 2024 review:

**The hillslope gully erosion mentioned on this line is for the formation of hillslope gullies and rills through hillslope erosion. We are not talking about mass movement or mass wasting. No mention of landslides or debris flows is provided in the manuscript in terms of modeling or empirical observations. There are plenty of examples of gully erosion definitions in various government reports from agencies on multiple continents, this manuscript subscribes to the common definition of the term and is discussed at length in the Discussions section. Mentioning gully erosion as part of the hillslope erosion modeling is described in the Methods section.
The onset of gully erosion is interesting in this region due to the prevalence of gullying on hillslopes post-wildfire. The process is more fully explored in an earlier paper from the nearby Schultz Fire that is cited already in this manuscript. The reference for that paper is as follows:**

Neary, D.G., Koestner, K.A., Youberg, A. and Koestner, P.E.: Post-fire rill and gully formation, Schultz Fire 2010, Arizona, USA. Geoderma, 191, pp.97-104, 2012.

Another reference to the Neary et al. study has been added at this line location.