Note to editor/reviewers. Plain text represents the original comment, and bold text represents the response.

Reviewer 3

This study uses a unique debris flow response dataset from a region with a paucity of post-fire debris flow information to test the USGS PFDF susceptibility and volume models commonly used in the western US for emergency management. The authors find preliminary support for the use of regional correction factors for the volume model and reveal potential drivers for reduced susceptibility 2 years following fire. Overall, the manuscript is clear in its objectives, well-written, well-supported, and presents important findings. I recommend some minor revisions to improve clarity and make some suggestions for presentation of results that the authors could consider.

Thank you!

Lines 116-117: Authors refer to Fig 1 when describing Quaternary-aged ("quaternary" should be capitalized, too) landslides but it is very difficult to see them in the hillshade as it is currently presented. Perhaps another figure in supplements if authors would like to show them, annotations directly on the plot highlighting the slides, or just no reference to Fig 1. Also, I don't know which watershed is Devil's Hole based on Fig 1.

Yes good point. I have now capitalized Quaternary, and I also removed the reference to figure 1. The way I had written it, a reader might have expected to see landslides mapped in figure 1. We don't have room for that, and this is really just an auxiliary background information, not a key point of the study. So I've removed the reference to figure 1 so that readers don't expect to see a figure that shows mapped landslides. To see that figure, they can follow the reference that I have in the sentence. Finally, to see the location of Devil's Hole watershed, see the outline of Figure 2c in Figure 2a.

Line 170: What are the models of tipping bucket rain gauges installed and their corresponding measurement resolutions?

I've now added this to Table 1.

Table 1. Rain gauges deployed in and around the Grizzly Creek burn area, operated by the USGS Colorado Water Science Center (USGS WSC), the Colorado Department of Transportation (CDOT), or the USGS Landslide Hazards Program (USGS LHP).

Rain Gauge Name	Owner	Station ID	Data Start	Data Stop	Data Gap	Rain Gauge Model/Tipping Bucket Depth (mm)
Cinnamon Creek Complex	USGS WSC	GCTC2	19 Jul. 2021	present	7/29/21 to 8/12/21	Vaisala WXT536/0.01
Cinnamon Creek	USGS WSC	GCCC2	19 Jul. 2021	present	No Gap	Vaisala WXT536/0.01
Deadmans Creek	USGS WSC	GCDC2	14 Jul. 2021	present	7/22/21-7/26/21	Vaisala WXT536/0.01
No Name	USGS WSC	GCNC2	15 Jul. 2021	present	7/28/21 to 8/12/21	Vaisala WXT536/0.01
Windy Point	USGS WSC	GCIC2	12 Jul. 2021	present	No Gap	Vaisala WXT536/0.01

East Fork Dead Horse Creek	USGS WSC	GCEC2	13 Jul. 2021	present	No Gap	Vaisala WXT536/0.01
Coffee Pot	USGS WSC	GCFC2	13 Jul. 2021	present	No Gap	Vaisala WXT536/0.01
Bair Ranch	CDOT	N/a	30 Jun. 2021	present	No Gap	Vaisala RG13H/0.02
USGS_gc_1	USGS LHP	N/a	17 Sept. 2020	present	No Gap	HOBO RG3M/0.02
USGS_gc_2	USGS LHP	N/a	17 Sept. 2020	present	No Gap	HOBO RG3M/0.02
USGS_gc_3	USGS LHP	N/a	17 Sept. 2020	present	No Gap	HOBO RG3M/0.02

Line 197: Shouldn't 4 km^2 be 4 km (distance not area)?

Good catch. Changed to 4 km.

Line 206: "see section 0" – there is no section 0?

Thanks for pointing that out. That was an automated reference that failed. I have now changed it to section 3.4.2.

Line 224: Were channel polygons hand-drawn or automatically extracted using a buffer around a flowline? Could be a nice detail to include.

I clarified that they were mapped by hand at the end of this sentence: "For each debris-flow observation in our inventory, we mapped erosional and depositional areas in each channel with separate polygons by hand."

Section 3.6: Recalculated dNBR post-recovery is a great idea and fits nicely with recent literature on quantifying vegetation recovery and its influence on debris flow susceptibility (such as Graber et al., 2023, link here: https://doi.org/10.1029/2023GL105101).

Yes many of the authors of that study, were co-authors on this study and some of the ideas espoused in that paper, originated from the work done in this manuscript.

Line 334: Was there evidence to support sediment exhaustion of the channels such as downcutting to bedrock? Could be good to include.

We saw that in some places. Here's a photo, you can see the layered bedrock in the channel. This has been added to the supplement.

Figure S8. Photo from the upper portion of the canyon showing channelized debris-flow erosion down to bedrock indicating sediment depletion.

Line 339: Not sure about using the term “nucleate” here and elsewhere (e.g. line 412) when referring to erosion/deposition in this context. I usually think of nucleation as a process that begins at one point and propagates outwardly, which I don’t think describes what’s happening here quite correctly. Could rephrase this as “initiate” or similar.

Changed “nucleate” to “initiate” in both locations.

Line 354-355: “changes between the debris flows and the lidar flight.” Clarify. Do authors mean to say: "changes to the debris fans occurring between initial deposition and subsequent lidar flights" or something like this?

Thanks, I changed the sentence to:

“The depositional volumes observed were less than the erosional volumes, as was expected due to sediment disturbance between the time of debris-flow deposition and the lidar flight (Figure 7).”

Line 360: How were Coal Seam and South Canyon debris flow volumes estimated? Just curious if this could exert some uncertainty in a comparison of these earlier datasets to the lidar/fan based estimates for the Grizzly Creek PFDFs. It is promising that they roughly show similar area-volume scaling as the authors point out.

Good question. The measurements for the Coal Seam and South Canyon debris flows were performed using the methods explained in Santi et al., 2008. Researchers made measurements in the field at cross-sections in channels measuring channel scour. I have added this sentence to the methods section:

“The volume data for the South Canyon and Coal Seam Fires were collected using the methods described by Santi et al. (2008) where researchers made measurements within channels estimating scour depth. The uncertainty differences between these field measurements and the lidar data are unclear; however, we estimate that the field measurements may be of a similar magnitude as the lidar (tens of centimeters).”

Fig 8B: There were no debris flows produced in 2022, correct? Maybe add in this language to figure caption since it is a bit ambiguous as is (no red stars = none correct?).

Added this sentence to the end of the caption:

“Note that no debris flows were observed in 2022.”

Fig 9A: The 1e5 scientific notation next to axis labels is too small. Consider blowing up this text or adding it alongside the units (e.g., 1e5 m^3).

I increased the fontsize for clarity. See new version of figure below in response to next question.

Fig 9B: I think the power law fits could be better coordinated with their respective point grouping colors – why have a Vp blue power law fit that does not match corresponding points with open red x’s and Vo black dashed fit that does not match red circles. Also, these power law fits do not visually seem to fit their respective datasets very well. Additionally, it would be good to provide an estimate of goodness of fit metric (R^2) as well as p-values (or confidence intervals) for regression parameters (prefactor and exponent) to provide some degree of confidence of these fits.

Some very good suggestions here. First, regarding colors, we agree, this can be improved. I think we have now simplified the colors to make it more clear. Here our revised logic for the new color choices. We are only fitting a line to the volumes predicted and observed at Grizzly Creek. We are then

superimposing the Coal Seam and South Canyon data points to show how they compare. Therefore, because we use the red/pink color for Grizzly Creek data, we've made the power-law fit lines red/pink as well. We just use different line styles to differentiate them. We have also changed the color of the South Canyon Fire data to yellow from black, so that it isn't confused as having a relationship with the lines in Figure 9a.

As for powerlaw fit. We agree that the fit isn't great, and that the data don't conform to the assumptions of a normal distribution. Therefore, reporting a p-value doesn't really make sense. However, we do report the R^2 values on the plot now (see below) and show +/- 2 standard error, which should capture 95% of the data.

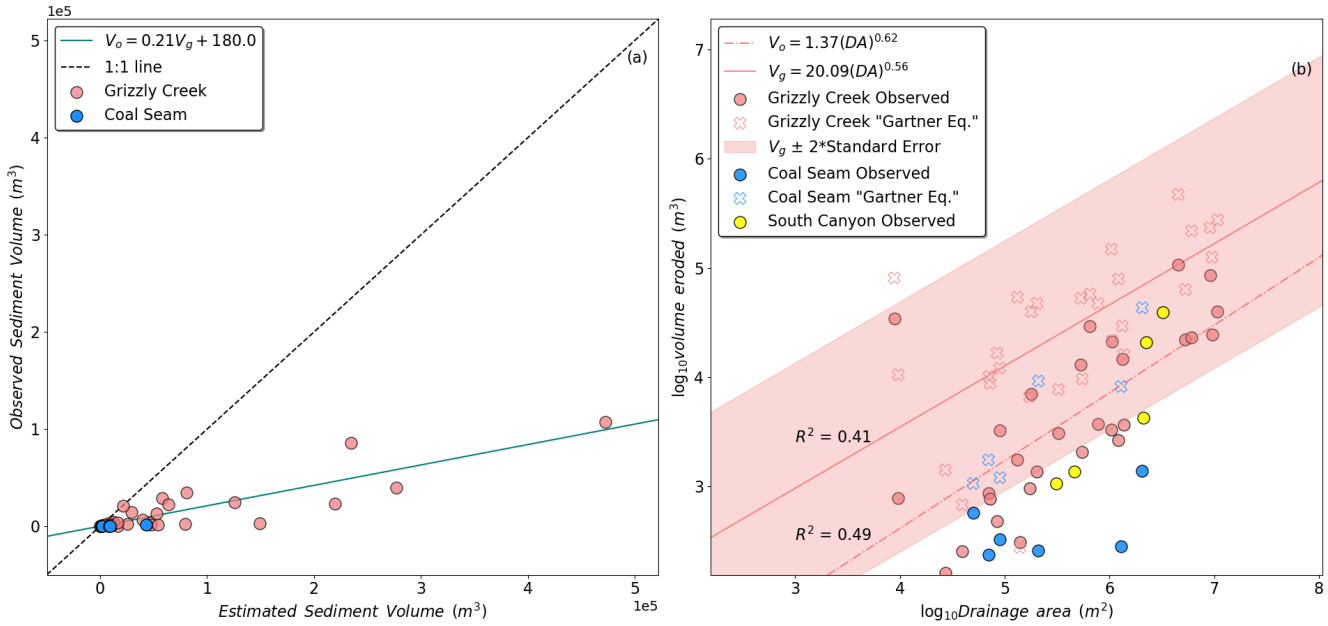

Figure 1. (a) Observed sediment volume (erosional) versus estimated sediment volume using Error! Reference source not found. **developed by Gartner et al. (2014). Linear trendline shows the relationship between the estimated (V_g) and observed (V_o) volume at the Grizzly Creek Fire. (b) A comparison of the total volume of observed sediment with estimated volume from** Error! Reference source not found. **as a function of upstream drainage area (DA). The observation volumes represent the volume of erosion, upstream of a transition to deposition. Best-fit power law equations were fit to the observed and estimated data points for the Grizzly Creek Fire, respectively.**

Additionally, for Fig 9B and the comparison between Coal Seam observed vs predicted (Line 360 earlier), do you see a similar ~4-fold overprediction from the Gartner et al. (2014) model? If it was close to this value, it further supports using this as a regional correction factor.

Because of the incoming data for the South Canyon Fire, we don't have sufficient information to generate a volume estimate using the Gartner equation, which we state with this text:

"Limited rainfall data for the South Canyon Fire precluded the use of Error! Reference source not found.**,…"**

Consequently, we can't say for certain the magnitude of the overprediction without applying Equation 2. However, we clearly see that the South Canyon observations are in-line with the observations made at the Grizzly Creek Fire. For the Coal Seam Fire, overpredictions range from 1.8x to 35x compared to the observations. Of the 6 observations, three of the overpredictions from the Gartner equation are <7.3x, and the other three observations are between 29x-35x. To highlight these differences to readers, we have added the following sentence to the discussion:
"The Coal Seam debris-flow volumes were overestimated between 1.8-35 times by Equation 2."