

Bold text indicates the original reviewer comment, and plain text indicates the response.

**Note to the editor:** When responding to a reviewer comment below, we identified an error in the data-limited dataset used to generate Figure 5 and the “Data-Limited Regions” values in Tables 5 and 6. The dataset inadvertently included a volume from northern California that we previously removed from the full volume dataset due to questionable rainfall data associated with the event. We removed this data point and updated Figure 5, Table 5, and Table 6, accordingly. As a result, there are several minor changes to the manuscript that were not directly prompted by the reviewer comments.

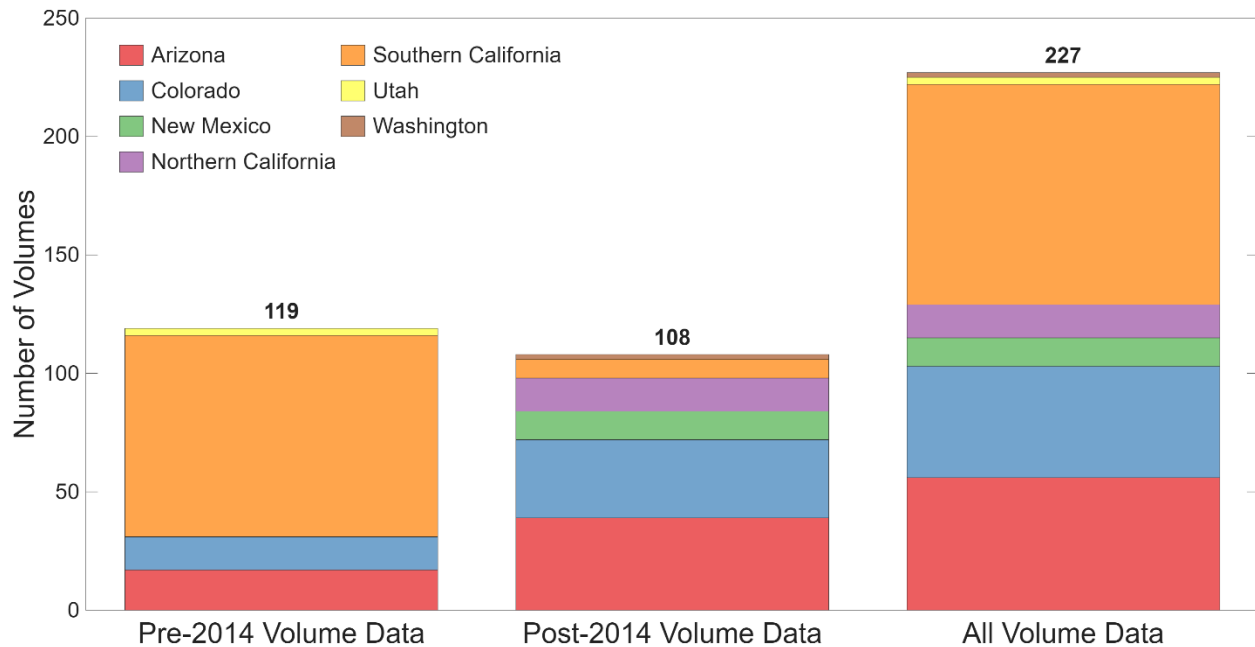
### **Editor**

**Thank you for providing your initial responses to the reviewers’ comments. Both reviewers have recommended minor revisions, and based on their reports and my own assessment, I concur.**

**I have two additional minor suggestions for your consideration:**

- 1. For Figure 6, please assess whether pie charts are the most suitable visualization. Because comparisons across categories are often easier with standard bar charts (i.e. avoiding the polar coordinate system of pie charts), a bar chart may improve interpretability. Please feel free to retain the current approach if you consider it more appropriate for your context; I simply wanted to flag this for your consideration.**

Thank you for this comment. After consideration, we agree that a bar chart may provide better interpretability for readers. We have updated Figure 6 so that it now includes three stacked bar charts that show the geographic distributions of volume collected prior to 2014, volumes collected after 2014, and all volumes included in the database used to develop the WEST model.



**Figure 6: Geographic distributions of the volume data used to develop the western United States (WEST) model, separated by date of volume measurement. Specifically, the distributions of volume data collected prior to 2014, volume data collected after 2014, and all volume data included in the database used to develop the WEST model.**

2. **For the sake of reproducibility, could you please add a brief description of the tools used for data (pre-)processing and modeling? There is no need to provide implementation details, but references to the main software packages and versions would be very helpful (e.g., R or Python versions and key libraries). For instance, your reference to ArcGIS Pro 3.3.0 for deriving terrain variables is sufficient on that front. In addition, we encourage all authors to share the code underlying their analyses where feasible - please consider whether this might be possible in your case.**

We used the Statistics and Machine Learning Toolbox in MATLAB R2024b to do our multiple linear regression analysis. We have added the following text to Section 3.2.2 of the methods:

“We then fit a multiple linear regression model to each combination using the Statistics and Machine Learning Toolbox in MATLAB R2024b, resulting in 702 unique, three-variable models.”

Because MATLAB is proprietary software, we are unable to share the code associated with our analysis. However, readers should be able to reproduce our workflow using a software of their choice using the methods outlined in the manuscript.

## RC1

### General comments:

**This study presents the WEST volume model, an updated postfire debris-flow volume model that uses an expanded dataset, including events from northern California and Washington. The model also introduces a rainfall ratio variable that normalizes rainfall intensity by the 1-year recurrence interval, improving model performance and transferability across diverse geographic regions. The model evaluation is extensive, assessing performance against the full volume database, regional subsets, prior models, and data-limited regions across the western United States, with performance metrics clearly presented and effectively illustrated. I particularly liked how the authors considered the prior sediment volume models and made a careful effort to build upon the previous work. The improvement in predicting sediment volumes is a significant contribution for postfire hazard assessments. Below I have listed comments that I think can be easily addressed.**

### Specific comments:

- 1. Regarding the final model selection, I noticed that models #1 and #3 in Table S3 include the same predictor variables with identical performance metrics, differing only in the use of the i30 vs. i60 rainfall ratio. Table 2 also indicates that both rainfall ratios have the same correlation coefficient (-0.13). Consider clarifying whether the selection of the i30 model was primarily based upon its sub-hourly duration. Additionally, how should practitioners proceed in situations where i30 data are unavailable, or where i60 may be more appropriate for operational forecasting applications?**

Thank you for this suggestion. We primarily selected the i30 model based on the criteria outlined in Section 3.2.2. In addition to having the best performance metrics, it used predictor variables that were commonly selected by the 29 final models, all three predictor variables had p-values  $< 0.1$ , and all three predictor variables had VIF values  $< 10$ . The difference in performance between the i30 model and other similarly performing models, however, was marginal, as noted. As a result, it is possible to use one of the alternative models listed in the supplement instead of the WEST model without sacrificing model performance. In that sense, it is reasonable to select the i60 volume model as an alternative if i30 rainfall data is not available.

We added the following text to Section 4.1 of the Results where we introduce the WEST model to alert readers to the fact that there are multiple models with similar performance metrics:

“Note that although we selected the WEST model as the final model for this study using the criteria outlined above, many of the final 29 models offer similar performance metrics (Table S3) and may be viable alternatives to the WEST model in some scenarios.”

We also added the following text to Section 5.2.2 of the Discussion, where we discuss potential use cases for alternative models that use *i15* or *i60* rainfall ratio, instead of *i30*:

“Finally, although we selected the WEST model using the criteria outlined in Section 3.2.2, several alternative models that we developed as part of this study offer similar performance metrics (Table S3) and may be preferred to the WEST model in some situations. For instance, we determined that rainfall ratio calculated over a 30-minute duration yielded the best results for our dataset, but there may be scenarios where models that incorporate rainfall ratio calculated over 15 or 60-minute durations are more suitable. It may be more practical, for example, to use a model that includes *i60* rainfall ratio, instead of *i30* rainfall ratio, to estimate volumes for a mitigation project if only *i60* design storms are available. Alternatively, it may be easier to implement a model that uses *i15* rainfall ratio within a hazard assessment framework that also predicts postfire debris-flow likelihood using rainfall characteristics measured over a 15-minute duration (Landslide Hazards Program, 2018). Given their potential applicability in these scenarios, we present alternative models that use *i15* and *i60* rainfall ratio, along with the same terrain and fire variables used by the WEST model (Equation 2), to predict postfire debris-flow volume in Table 9. Note that, although it marginally outperforms the WEST model, the *i15* model does not pass the Anderson-Darling test ( $p$ -value  $< 0.05$ ), so was not considered for final model selection. The *i60* model, on the other hand, performs similarly to the WEST model and meets all requirements of multiple linear regression (Table 9).”

Finally, we added the following table to Section 5.2.2 of the Discussion that shows a comparison between the WEST model and alternative volume models that use *i15* and *i60* rainfall ratios:

**Table 9:** Equations for volume models that use rainfall ratios calculated at different durations, including 15 minutes ( $i15_{rr}$ ), 30 minutes ( $i30_{rr}$ ), and 60 minutes ( $i60_{rr}$ ). All models predict debris-flow volume ( $V$ ) as a function of rainfall ratio, watershed area ( $a$ ), and watershed area burned at moderate or high severity with slopes  $\geq 50\%$  ( $mh_{50}$ ). Performance metrics include  $R^2$  and root mean square error (RMSE), as well as the  $p$ -values for the Anderson-Darling (AD) and Brown-Forsythe (BF) tests.

ID	Model	$R^2$	RMSE	AD Test p-value	BF Test p-value
<i>i15</i>	$\ln V = 7.82 + 0.35 \ln i15_{rr} + 0.76 \ln a + 1.09\sqrt{mh_{50}}$	0.67	1.30	0.03	0.17
WEST	$\ln V = 7.56 + 0.20i30_{rr} + 0.75 \ln a + 1.11\sqrt{mh_{50}}$	0.66	1.31	0.07	0.17

2. Consider including Gatwood et al. (2000) in your statement on line 79, “There are several postfire debris-flow volume models that include rainfall variables,” especially as Gatwood et al. (2000) uses the maximum 1-hour precipitation.

We agree that Gatwood et al. (2000) is a useful reference and have added it where suggested.

3. I think that the statement surrounding the standard deviation on line 499 is not in agreement with the standard deviation values listed in Table 6. The in-text statement reads “The WEST model slightly overpredicted volumes from data-limited regions but had the lowest standard deviation of the four models (Table 6).” I believe Table 6 lists the lowest standard deviation from the data-limited regions as 1.29 (EAV model) not the WEST model (1.37).

Thank you for catching this discrepancy. The data in Table 6 is correct; the residuals of the EAV model had a lower standard deviation than those of the WEST model. We have updated the text so that it now accurately reflects this information. The updated text is:

“The WEST model slightly overpredicted volumes from data-limited regions but had the second lowest standard deviation of the four models (Table 6). The EAV model had a slightly lower standard deviation than the WEST model but overpredicted volumes from data-limited regions more substantially (Figure 5b). Unlike the WEST and EAV models, the V1 model underpredicted volumes from data-limited regions, on average (Figure 5d).”

4. “Rainfall ratio” and “rainfall anomaly” are used interchangeably throughout the manuscript and the supplemental. For example, on line 415, “rr” is used in Eq. 2, whereas “ra” is used in Table S3. Although everything is clearly defined, maintaining consistent naming of this variable throughout the manuscript and supplemental may improve clarity for the reader.

Thank you for pointing out this inconsistency. We have updated the manuscript and supplement so that we now only use the term “rainfall ratio” in the text and “rr” in equations.

5. Although Table 1 displays the number of volume measurements geographically, it lacks information surrounding the magnitude (e.g.,  $10^1 - 10^4$ ) of the sediment volumes geographically. Consider adding a column with the range of sediment volumes in Table 1.

We agree that volume magnitude is an important piece of information that is currently missing from our tables. To rectify this, we have added a column to Table 1 titled “Debris-Flow Volume” that includes the minimum and maximum debris-flow volumes associated with each site.

- 6. While the need for additional volume data from data-limited regions is discussed, it may be helpful to also acknowledge the importance of including larger-magnitude sediment volumes, as the current dataset from these regions contains smaller volumes.**

This is an important point, so we have added the following text to Section 5.2 of the Discussion:

“In particular, volume data from larger-magnitude debris flows in data-limited regions are needed to more fully evaluate the performance of the WEST model in these settings, as the median volume of the 19 debris flows from data-limited regions included in this study was 511 m<sup>3</sup> (Table S2), nearly five times lower than the median volume of 2,550 m<sup>3</sup> associated with the entire volume database.”

#### **Technical Corrections:**

- Table 1 – consider adding more text for context to the Table 1 caption (e.g., “Fire information associated with measured sediment volumes”)**

We have updated the caption of Table 1 to provide more information about the data contained in the table. The new caption is:

“Table 1: Fire and associated debris-flow information. Fire information includes the fire name, year of occurrence, and location, including Arizona (AZ), California (CA), Colorado (CO), New Mexico (NM), Utah (UT), and Washington (WA). Debris-flow information includes the number of volume measurements for each fire, the range of associated sediment volumes, and the original sources of the volume measurements.”

- Line 123, add space after “,” in “Smith et al.,2021”**

We have added the space where suggested.

- Figure 1 – consider simplifying the fire icon to a circle symbol for improved legibility and/or lighten the ecoregion colors to increase contrast between the ecoregion and the volume symbol.**

We have updated Figure 1 to improve the contrast between the ecoregions and volume symbols. Specifically, we adjusted the color scheme of the volume symbols to better distinguish between the volume categories, as well as between the volume symbols and ecoregions. We also updated the basemap to improve overall readability.

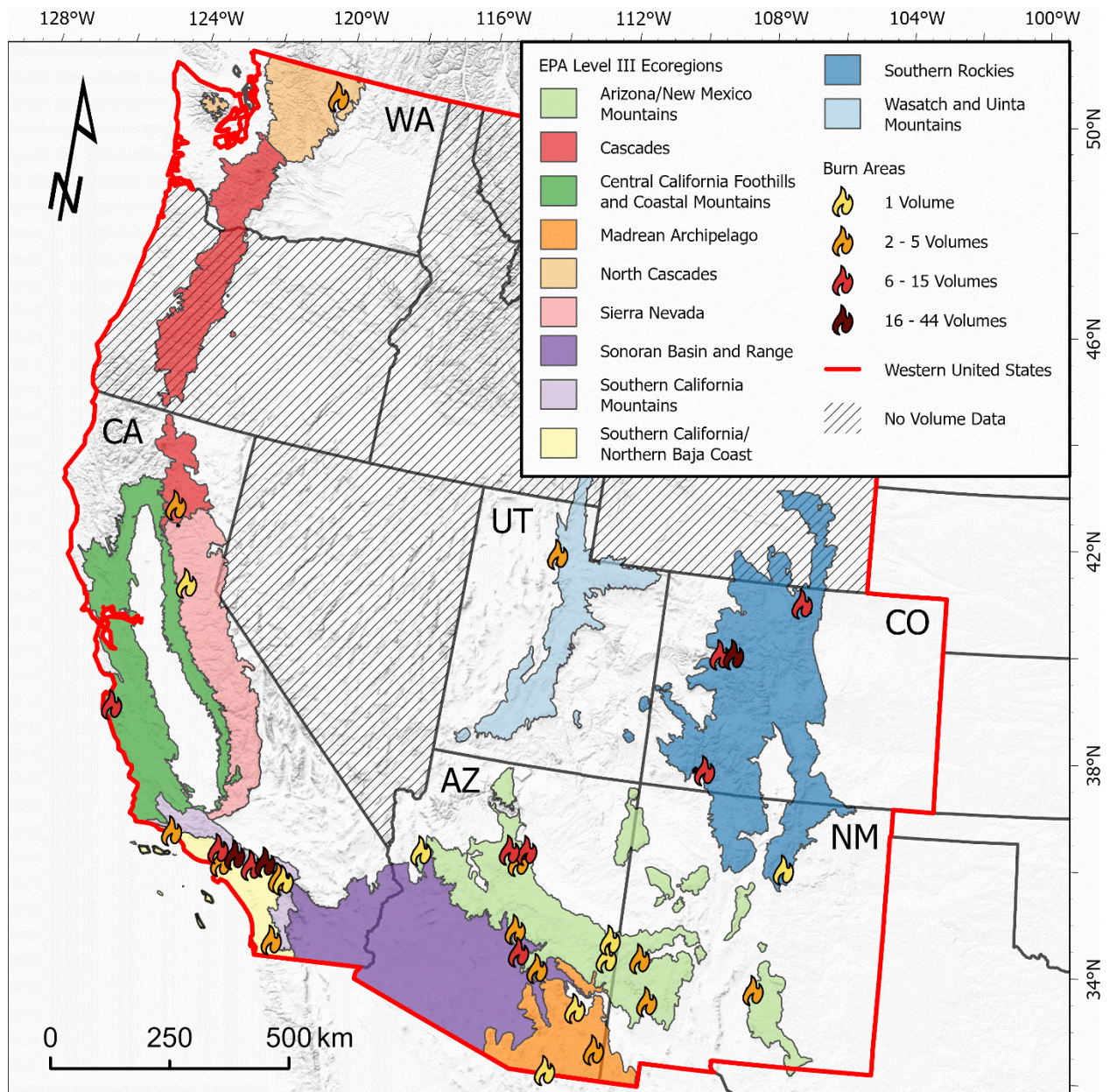


Figure 1: Map of the locations of the 34 burn areas included in this study. The burn areas span six states across the western United States (US), including Arizona (AZ), California (CA), Colorado (CO), New Mexico (NM), Utah (UT), and Washington (WA), and 11 Environmental Protection Agency (EPA) Level III Ecoregions. The names of the ecoregions shown in this figure are derived directly from the EPA (U.S. Environmental Protection Agency, 2013). Basemap credits: United States Geological Survey The National Map: 3D Elevation Program, United States Geological Survey Earth Resources Observation & Science Center: GMTED2010.

- Line 256 – for increased clarity, consider including the text “for each watershed” in the phrase “We calculated another 17 variables for each watershed related to fire characteristics (Table 4)...”

We have added the phrase “for each watershed” where suggested.

- **Line 271 – delete “and” directly after “B<sub>k</sub>”**

We have deleted “and” where suggested.

- **Line 415 – consider spelling out the first instance of “VIF” in this section for clarity.**

We redefined VIF at first mention in the results section, as suggested.

- **Lines 457, 462, and 523 – change “Figure S2” to “Figure S3”**

We have updated these references from “Figure S2” to “Figure S3.”

- **Line 460 – delete “a” directly after “MAE (Table 7) and”**

We removed “a” where suggested.

- **Line 489 – change “Figure S5” to “Figure S5a” to maintain consistency with “Figure S5d”**

We changed this reference to “Figure S5a” as suggested.

- **Line 657 – change “Figure 6” to “Figure 7”**

We changed this reference from “Figure 6” to “Figure 7.”

## **RC2**

### **General comments:**

**Thanks for the opportunity to review this MS. The MS describes the development of an empirical model of post fire debris flow volume that draws on a greater quantity of observation data and aims to create a more general model for the Western US than the current models developed from datasets collected in different regions. The paper is well suited to the audience of NHESS. The overall objective is sound ( a more generic model), The methods are well suited to the aim and are executed in an extremely structured and well defended and explained way. I particularly liked the extensive detail justifying the logic of the decisions regarding the acceptance/exclusion of the many possible models. The performance comparisons with existing models were really well justified in terms of the metrics used but also the graphics and tables. The authors conclusions and interpretation of the results were clear and concise. Overall this is a very high quality MS. I only have two minor suggestions.**

### **Specific comments:**

- 1. One relates to the figures of distributions of the residuals; would it be possible to include a second x axis with the untransformed values for this distribution, as it is difficult to interpret in the context of units that are intuitive.**

Thank you for this comment. Because all four volume models were developed in natural logarithmic space, the residuals represent differences in log space. Transforming the residuals into dimensional space would change their distribution. As a result, a second x-axis with untransformed (dimensional) values would not accurately reflect true values of the residuals if they were transformed into dimensional space.

However, we appreciate that the current x-axis values in these figures are not the most intuitive. To improve the interpretability of these figures, we have instead added a second x-axis that shows the ratio of the predicted volume to the observed volume. We calculated this ratio by exponentiating the natural logarithmic residuals. For example, if the residual for a given volume is 0, indicating no difference between predicted and observed volume, then the ratio of predicted to observed volume would be 1, because  $e^0 = 1$ .

We added a second x-axis that shows the ratio of predicted to observed volume to Figures 3 and 5 in the main text and Figures S3, S4, and S5 in the supplement.

# Western United States Predicted / Observed Volume

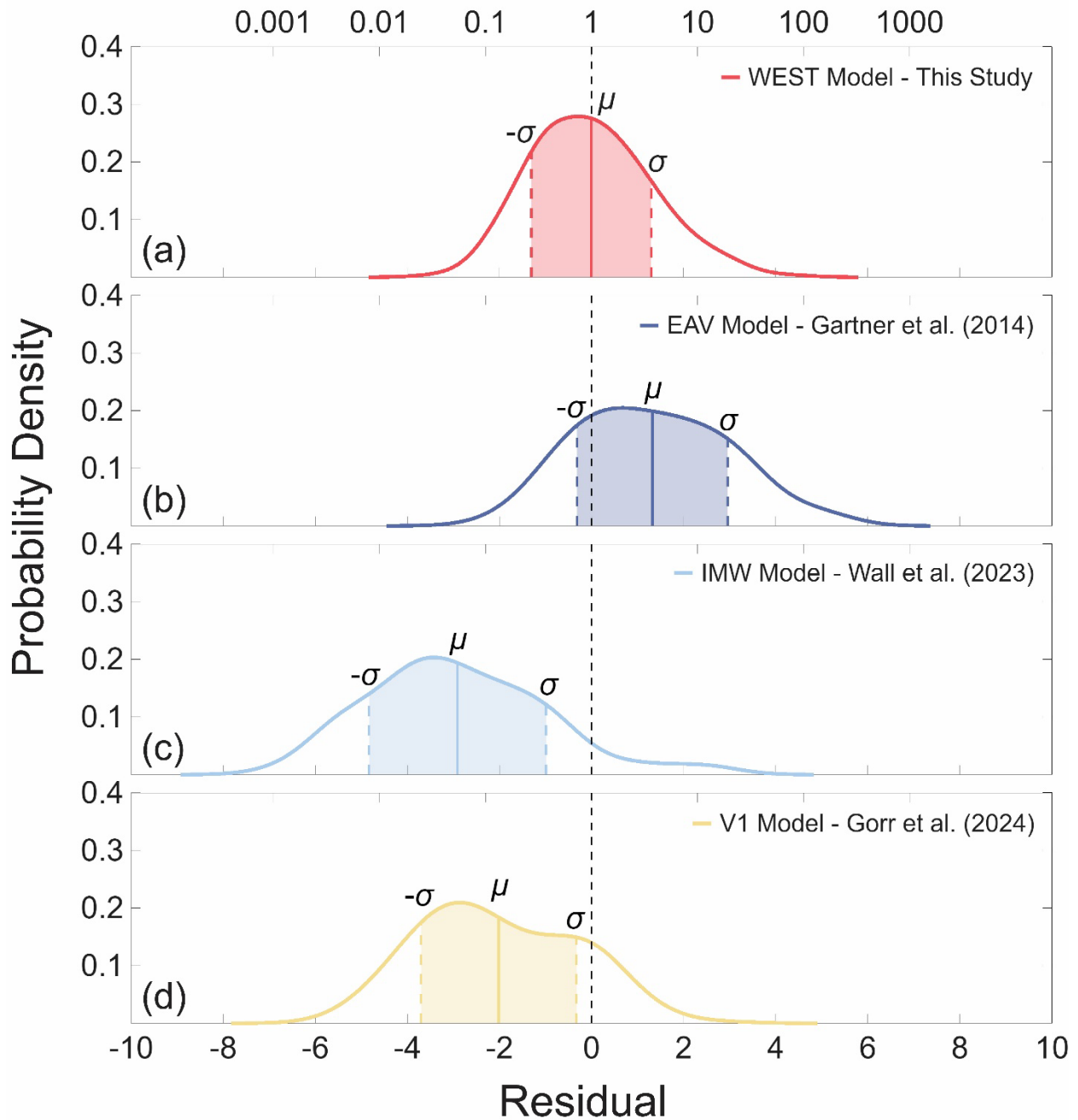


Figure 3: Probability density functions for the residuals of the (a) western United States (WEST), (b) Emergency Assessment volume (EAV), (c) Intermountain West (IMW), and (d) V1 models when applied to the entire western United States dataset.

- 2. The second point is, I know the scope of the model development is restricted to the western US, but it would be good to see the results interpreted a little in the context of the rest of the world. How does this model/modelling approach/variables compare to what others have published on this topic? What might the implications be for this new analysis for other researchers with similar objectives in other regions/continents re estimating post fire debris flow volumes? This could make the US work more relevant to a wider audience than western US practitioners and researchers.**

We are not aware of any postfire debris-flow volume models that have been developed for use outside of the western United States. However, findings from previous studies of postfire debris flows in Australia indicate that the primary controls on postfire debris-flow volume are broadly similar across geographic regions. We agree that discussing the implications for international researchers is critical, so we have added the following paragraph to Section 5.2 in the Discussion:

“Similarly, more data are needed to evaluate the performance of the WEST model when applied outside of the western United States. Although postfire debris flows are common hazards in fire-prone regions around the world, including Australia (e.g., Nyman et al., 2011), Canada (Hancock and Włodarczyk, 2025), Italy (e.g., Esposito et al., 2023), and Spain (García-Ruiz et al., 2013), debris-flow volumes and associated rainfall data remain scarce outside of the western United States. This lack of data has limited the development of postfire debris-flow volume models for these regions and has prevented the evaluation of most volume models developed for use in the western United States, including the WEST model, when applied elsewhere. However, previous studies suggest that the primary controls on postfire debris-flow volume may be consistent across geographic regions. For example, Nyman et al. (2015) found that watershed area was the most important control on the volume of 10 postfire debris flows that initiated in southeast Australia and that volumes could be accurately predicted using an empirical volume model developed for use in the western United States (Gartner et al., 2008). Similar to the WEST model, this model predicts postfire debris-flow volume as a function of rainfall characteristics, watershed area and slope, and burn severity (Gartner et al., 2008). These similarities indicate that the primary controls on postfire debris-flow volume are generally transferrable across geographic settings, although additional volume data from fire-prone regions around the world are needed to evaluate the performance of the WEST model beyond the western United States.”