Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC3
https://doi.org/10.5194/nhess-2021-49-AC3, 2021

Reply on RC2

Stuart R. Mead et al.

Author comment on "Quantifying location error to define uncertainty in volcanic mass flow hazard simulations" by Stuart R. Mead et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-49-AC3, 2021

Thank you for your review of our manuscript and suggestions, particularly regarding image composition. We have addressed your comments as follows, and have uploaded here (as supplement) revised images following your suggestions, as the most substantial changes.

1. Line 104 (Figure 1): This figure contains multiple sets of black dotted lines (crater outlines, debris flow deposit, contours, and spurious elevation zone), which is confusing to interpret. The craters are apparent and only need labeling (and perhaps arrows); the contours, deposit, and spurious zone should all be different designs and/or colors; and all of the labels need to be larger in order to be readable.

We have made changes to Figure 1, caption and text (see attached .pdf) based on these suggestions to improve interpretation.

2. Line 227 (Figure 2): Using a single color gradient rather than a rainbow here would make this figure much easier to interpret. Rainbows are subject to misinterpretation (see http://iis.seas.harvard.edu/papers/2011/borkin11-infoviz.pdf) and it's not easy to see the subtle variations in these simulations based on a rainbow colorscale.

We have changed Figure 2 to use a better color scheme for the sequential data, based on colorbrewer. This scheme is also consistent across all depth measures (Figure 3, categorical) as suggested in following comments.

3. Line 237 (Figure 3): There's no depth comparison going on here, only a comparison of inundation limits. Can you show the difference between the simulated and actual deposit depths as a gradient layer?

We are unable to reliably extract a difference between simulated and actual deposit depths due to the differences in post-event LiDAR which does not represent the immediate post-event morphology (see lines 109-127). The modelling approaches also do not explicitly consider stopping of material (see line 239), which would make quantitative comparisons suspect. This limits the analysis to a qualitative assessment of inundation limits. Around line 240, we have modified text to highlight the comparison is between deposit outlines ("The predicted deposit **outlines** for all simulations …")

3b. Additionally, this is a different gradient than in the previous image. It's better (see *comment about rainbows) but it would be nice to see some consistency in colors depending on the variable being shown (depth, PPV, etc.) Also, the outlined/white dotted zone at the bottom right needs a label, assuming it is the source zone for the simulations.*

Colorscales are now similar for heights (Fig. 2 & 3; note one is continuous vs. categorised) and PPV is a separate color gradient (Fig. 6).

4. Line 276 (Table 2): Why do the basal & internal friction angles and the solid volume fraction parameters vary between these two models? In a first-order comparison of their ability to accurately represent a flow deposit, wouldn't you want to make sure that all the shared parameters are consistent between the different models? If not, the reason for these choices needs to be documented in the methods section.

(we presume you mean ~Line 219/Table 1) The models themselves vary in completeness of the model physics or take a rheological approach, which result in a difference between shared parameters. The large difference between Voellmy and two-phase model basal friction is due to the application of basal friction to the solid fraction only. The calibration approach (as in McDougall 2016) may also result in small differences between the models, but the larger part of the difference is due to the different models applied. This has been clarified in text ("*Best-fit values for similar parameters (basal friction and solid volume fraction) in Table 1 vary as a result of the differing drag contributions, parameter sensitivity, rheological and constitutive models and the calibration approach. For example, the two-phase models only apply basal friction to the solid volume fraction, whereas the (single phase) Voellmy-Salm approach considers a bulk basal friction of the fluid-solid mixture, and additional viscous stresses in the Pudasaini (2012) model appeared to reduce the sensitivity and value of basal friction.")*

5. Line 283 (Figure 6): The debris flow outline is difficult to see here. Perhaps a colorscale that's less saturated on the high end, or a diverging one?

We have applied a less-saturated version of the PPV colorscale.

6. Line 321-322: What are 'reasonable' values? What is reasonable to one practitioner may not be for another, or in a different setting.

Reasonable was an unsuitable term here, we have revised the text to clarify the values are within the range of those identified in previous literature.

7. Line 336 (Figure 8): How did you choose 0.25 as your fuzzy quantity limit? What makes this value important from a hazard zonation standpoint? Additionally, the figure appears to show hazard zones overlain on the debris flow deposit outline, but there's no label distinguishing the two outlines or pointing out the source area (assuming it is the cross-hatched zone at the upper end of the flow).

The fuzzy quantity has a similar effect to length scale, but has more variability at small length scales due to the discrete approximation of a gaussian. We chose to set the quantity to 0.25 as the smallest value possible in a 3-cell weighting kernel. This needed to be clarified, so we have modified Figure 7 to show fuzzy quantities of 0.1, 0.25 and 0.5, and added a short explanation to the end of section 4.1.
Figure 8 has been modified to add labels.

8. Line 337: Is the 'length scale' being referred to here the length scale of the correlation function? You may wish to specify so as to avoid confusion.

Yes, length scale of the correlation function (equation 2). Modified in text.

9. Line 345: When you say "map scale", do you mean the resolution of the model results, or the DTM? Can you explain how you derive the map scale, since the reference you include is old and not available online? It may seem simplistic to an experienced practitioner, but going into more detail will mean that other researchers will be able to more easily apply your methods.

“Map scale” here refers to the scale of the hazard map (i.e. including all features it may be important to include in a hazard map for communication). We have modified wording to clear this up. We have also added a short explanation for how the map scale denominator is obtained (by multiplying length scale by 1000) from the Tobler method.

10. Line 350-352: I find this wording confusing. Why else would you put the effort into evaluating the output of models if they aren't going to be used to generate hazard zones in the long run? Do you mean instead, "this method is useful for quantifying the uncertainty of modeling output, which can be assist in generating hazard zonations according to specific risk tolerances, but does not produce raw data appropriate for public products"? Lines 367-370 are a less confusing way of stating this, and would do better in place of the current text.

This (lack of a mathematical basis to parameterise length scale) is a limitation of our method as currently defined, we think this limitation should be communicated largely as written. To reduce wording confusion, we have made minor modifications to text on these lines to emphasise it is a **current** limitation that could be alleviated and the method cannot be used as an **automated** process to generate hazard zones for uninformed stakeholders (i.e. ‘public’).

11. Line 363: Worded confusingly - break this up into a couple of clear statements to drive the point home.

Broken up into a clear statement and example. ("We found increasing the length scale (λ) of the correlation function can increase performance metrics for a commensurate decrease in the opposing performance metric and resolution. For example, an increase in sensitivity will result in a decrease in positive predictive value.")

In addition, all text errors have been addressed.

Please also note the supplement to this comment:
https://nhess.copernicus.org/preprints/nhess-2021-49/nhess-2021-49-AC3-supplement.pdf