# Reviewer 2:

This manuscript discusses mapping mass-movements in alpine environments in SAR interferograms. They focus on two aspects: the first being the variability among human experts mass-movement phase patterns and the second being the automatic mapping of the same interferograms by several deep neural networks, including a modification of the well-known UNet++ architecture (Attention UNet++). For this purpose, four Sentinel-1 interferograms with temporal baselines of 12 days (three) and 18 days (one) of the Valais canton in Switzerland were computed. They find a large variability in the human-mapped mass-movements of one interferogram ranging from 0.18 - 0.49 IoU. After evaluating various convolutional neural networks, UNet++ with the ResNet-18 encoder was chosen based on a very extensive sensitivity analysis. The performance of this network falls within the range of variability of the experts.

This manuscript is pertinent to land-deformation studies, natural-hazard research, and the broader InSAR data user base that intends to use deep learning for downstream segmentation tasks. While I appreciate and commend the authors on their methodological thoroughness, I see two major points that need to be addressed before the work can be published. First, this is a very long manuscript. I recognize the effort put into organizing it into many, many subsections. However, I think the length makes it conducive for a reader to lose the track(s) of the study (as I myself did several times). For this reason, I would recommend keeping the sensitivity analysis out of the main manuscript. Second, I agree with the concerns raised by referee 1 (https://doi.org/10.5194/egusphere-2026-375-RC1) regarding the exclusion of negative samples from the training dataset and the weak justification for this in the discussion. In addition to these, there are a few specific comments and technical errors which also need to be addressed. I have listed them below.

We thank the reviewer for the time taken to evaluate our manuscript and the constructive comments. In the following, we respond to each comment individually. We acknowledge the length of the manuscript has become quite extensive and will follow the reviewer's advice to exclude the sensitivity study from the main text and move the corresponding content (incl. current Figure 7 and Table 4) to the appendix.

Specific comments

1. Line 182: How do make this assumption when it is impossible to know the impact of training a neural network in a supervised way with "missing labels"? What do you mean to convey here?

We thank the reviewer for this comment and agree that the original wording is ambiguous.

Our intention was not to suggest that the relative impact of missing versus incorrect labels can be rigorously quantified, but rather to describe a practical design choice made during training. Due to the strong class imbalance and the use of extensive data augmentation, annotated positive samples are seen repeatedly by the model, whereas background signals are predominantly learned implicitly at the pixel level. In this setting, incorrectly labelled positives (i.e. false positives) are reinforced multiple times during training and may therefore have a disproportionate influence on the learned representation. In contrast, missing labels (i.e. unlabeled positives) are treated as background and are not consistently reinforced in the same way. We therefore prioritised high confidence in the annotated signals, even at the expense of potentially omitting uncertain or borderline cases.

We agree that this reasoning was not clearly expressed in the manuscript and will revise the text accordingly to clarify this point.

2. Line 188: How do you still have training samples with velocities starting from 10 cm/a when they are supposedly filtered out as mentioned in line 186?

We thank the reviewer for catching this error! The statistics still included the unfiltered version. We correct the values to ~20 cm/a to ~210 cm/a in the revised version.

3. Line 357: You state that the full dataset consists of ’centre-based’ sampled patches while having stated earlier that both negative and positive patches were sampled with a sliding window (Section 2.3.4, lines 215 - 220) and that the full dataset contains both the positive and negative patches (lines 226 - 227). Are the negative data samples constructed with ’centre-based’ sampling? In Table 3 there is no mention of another dataset. The sentence ’Minor decreases compared to models trained on centre-based sampled data...’ implies the opposite of what the metrics show in Table 3. Please rephrase this sentence to make it clearer.

The sentence in question is referring to the ISSLIDE model, which was trained on centre-based data. We revise to “Minor decreases compared to models, such as the ISSLIDE model, which are trained on centre-based sampled data demonstrates that our approach is robust to artifacts in interferograms that could resemble mass movements (e.g. local atmospheric patterns, noise), making it suitable for effective application over large areas. “

4. Section 4.2.1 contains important information as well as repeating information already mentioned in Section 2.3.2 about the labels used to train the neural networks. Please consider merging the redundant points into Section 2.3.2.

Thanks, we will disentangle the redundant information and merge it into the suggested section.

5. Lines 500 - 504: Is the evaluation on the test set from Bralet et al., 2024 being alluded to in Section 3.2.1 as the ’centre-based’ sampled dataset? Why is it mentioned only in the discussion and not in the results? Also, simply stating that ’performance dropped by approximately 11%’ when running inference on a test dataset containing negative samples is vague. Which metric are you referring to? In a similar vein stating "When predicting over a complete raster, such as the test area in Queyras Park, the performance degrades even further" is not enough to support your claim. I agree with the comments

and suggestions from referee 1 regarding the exclusion of negative samples from your training dataset.

Yes, thanks for catching this ambiguity, we were referring to the IoU. However, after reexamining the ISSLIDE database, we found that we lack the outline of the actual mapped area and that the shapefile extent is not the extent mapped. Also the name-giving park outlines, while intersecting the mapped area, do not correspond to the actually mapped area either. We can therefore not provide meaningful information with false negative areas. As a result, we will fully take this statement out of the manuscript.

Concerning the exclusion of negative samples please refer to the answer given to reviewer 1.

6. Lines 514 - 517: I would also mention the other metrics used for assessing the semantic segmentation in Appendix E.

Thanks, for the sensitivity tests we made use of the IoU metric only, see table in appendix E. The other performed tests described in Tab. 3 were used to assess model performances. We acknowledge that this might be misunderstood, when not actively looking into the appendix section, and will make it more clear in the main text.

7. Lines 530 - 533: Could you please provide this information as a proportion/percentage of the total number of samples? Also, what is the slowest mass-movement captured by SAR that is detectable by a human mapper? How many such slow events actually occur in nature? Answering these questions could help in figuring whether it is even a problem that needs to be addressed.

Thanks for the questions. First, we understand that the original wording of these lines was ambiguous. In the manuscript, the relatively fast-moving processes are inferred to be rock glaciers and rockslides, whereas the slower-moving processes are inferred to represent larger, deep-seated slope deformations. We will adjust to clarify that we do not have these classifications but, because of the short temporal baselines used for labelling, our training set does not typically capture slower velocities (< 15 cm/a) that are attributed to larger, deep seated landslides. We will reformulate to:

> "At present, the model performs well in detecting relatively rapid mass movement processes (those with velocities between ~15 cm/a and ~210 cm/a which are typically associated with rock glaciers or rockslides). However, the model's performance for slower-moving processes, such as deep seated gravitational slope deformations, is likely limited. This is because deformation rates within this velocity range do not generally produce coherent fringe patterns in the 12 to 18 days S1 interferograms that were manually labelled and used for training."

The second question raised by the reviewer strongly depends on the imagery used during mapping. The detectable velocity is related to the temporal baselines and frequency ranges the human mapper is working with. As a general rule, fringe patterns of 0.25 are detectable in interferograms, but can often be unclear, when working on single interferograms, whether they are due to atmosphere or actual movements. Furthermore, the use of different colour scales within the community affects visibility, especially when the colour sensitivity of the human eye and the circular nature of the phase signal are not adequately considered, thereby further biasing detectability. This is why, in this work we trained the algorithm on detecting fringe patterns of minimal 0.5 fringe to minimise false-positive detections and minimise such biases. However, the resulting lower bound on velocity depends on the temporal baseline selected for the interferogram as well as on the signal wavelength – for our setup this corresponds to 15cm/a as used above.

On the third raised point: Large deep-seated landslides can move at comparatively low velocities over long periods before potentially accelerating, sometimes culminating in catastrophic slope failure (Agliardi et al., 2020). Such large slope instabilities occur less frequently in nature than smaller and more rapid mass movement processes, as described by the established power-law relationship between landslide magnitude and frequency (Stark and Hovius, 2001). Nevertheless, despite their lower abundance, these slow-moving landslides remain highly relevant, especially in alpine environments such as the Alps. They commonly affect large spatial extents, persist over long timescales, and may evolve into rapid failures (Agliardi et al., 2020). To further assess the representativeness of our mapped landslides, we performed a magnitude–frequency relationship comparison between our labels and a local inventory (Appendix A). This comparison provides an indication of the landslide sizes expected within the study region.

8. A recent review article published in this journal (https://doi.org/10.5194/nhess-26-487-2026) discusses the use of RNN's, LSTM as well as transformers, particularly to make use of the temporal dimension. You mentioned that it is standard practice for domain experts to use a time series to make sure what they delineate is actually a mass-movement (Lines 458 - 460). I would suggest adding a few sentences about this in either Section 4.2.3 or in Section 5.

Thanks for the suggestion, we now implement the new literature into the paper. Our suggestion for temporal evaluation would be to run the currently trained model as is over multiple interferograms and then analyse the fringe likelihood masks over time like this and discharging features that are not traceable in a systematic way over time. We will implement the suggestion in the text.

Technical comments

1. Line 24: This is a persnickety comment, but repeat acquisitions already implies that they are over the same area at different times.

Thanks, we delete "over the same area at different times".

2. Line 25-26: "multiple/multi-pass (ascending/descending) acquisitions and derives...". And I presume you mean displacement time series/ time series of displacement rates?

Thanks, we adjust to “by exploiting multi acquisitions and deriving time series of displacement rates.”

3. Line 28: Landslide state of activity.

Thanks, we add the “of”.

4. Line 29: It might be worth elaborating the factors/reasons for poor coherence in mountainous areas.

Thanks, we adapt to:
“However, PSI techniques require high temporal coherence on radar backscatter, which is rare in alpine environments due to changes in snow cover, soil moisture and vegetation (Karbou et al., 202) as well as fast moving surface processes (e.g. sediment transport (Olen et al., 2020) or fast moving landslides (Manconi, 2018)). This limits the performance in terms of spatial coverage and accuracy especially in rapidly moving areas Manconi, et al., 2021.”

5. Line 34: "...when PSI does not." is an incomplete sentence.

Thanks, we changed it  to “...when PSI cannot.”

6. Line 47: Add a comma before the citation and remove parenthesis around the citation.

Thanks, we adjust accordingly.

7. Line 55: Change the apostrophe in area and elevation numbers to a comma. Please make this change for other numbers in this paper, where applicable

We thank the reviewer for the suggestion, since the comma-style is mis-interpretable and not defined for in the guidelines from NHESS, we will adjust the writing style but make use of the SI preferred standard of a space to indicate thousands e.g. 10 000.

8. Figure 1: While I can appreciate the authors trying to convey a lot of information in a summarized way through this workflow figure, the subsection numbers above each process is distracting. I would recommend removing these, also because I mentioned in the specific comments that there are too many subsections and sub-subsections fragmenting the paper.

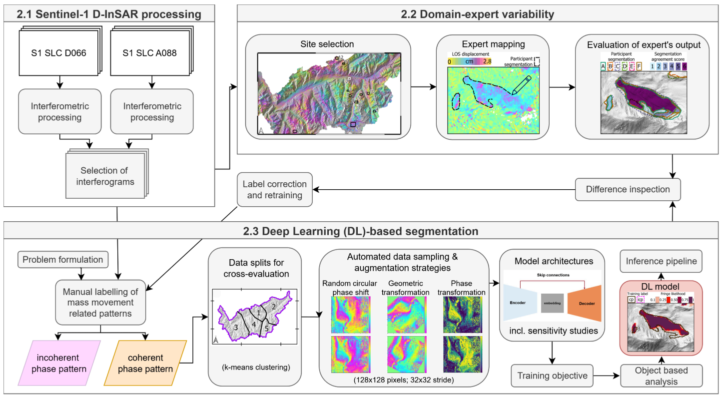
   Some more comments to this figure:

   (a) Please remake the figure with at least 300 dpi.

   (b) Please rename "Setup" in the Domain-expert variability block (and the corresponding subsection title) to something more descriptive. I would choose another interferogram to show the delineation part, maybe something without shadow?

(c) Please remove the sub-panel labels ’3a’ and ’b’ from Evaluation of expert’s output block.

We thank the reviewer for helping us make the plot more readable. We adjust accordingly:




9. Line 62-72: There is a switch between active and passive voice from sentence to sentence. Please stick to one voice.

Thanks, we adjust the entire paragraph to use the active voice.

10. Line 74: ’In the following subsections, ...’

Thanks, we add the “subsections” into the sentence.

11. Line 81: Sentinel-1 has 6 day repeat globally when two satellites are active.

Thanks, we remove the corresponding part.

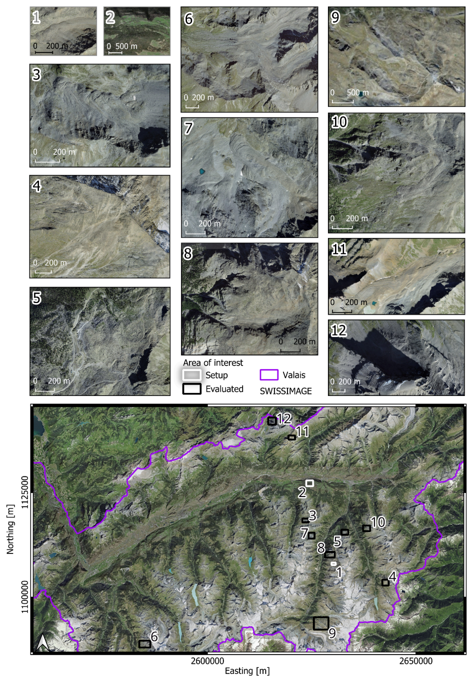
12. Figure 3:

(a) Please use a different color for the polygon indicating the "Setup" areas.

(b) I and perhaps other readers would be interested in knowing what exactly are the different patterns in these AOIs?

In lines 99-101 you make a distinction between sites 3-8 and 8-12. Maybe you could mention the different mass-movements as subtitles of the sub-panels?

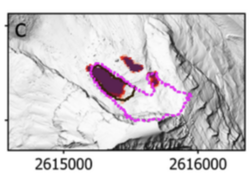
a) Thanks, we will make use of a colour which is better visible than the current gray.
b) The scenes are often more complex than just one process. We performed a process-classification study with seven landslide experts classifying the landslide processes that resulted in movement detected in the interferogram by the InSAR experts and the results were found multi-modal. So we deem giving one title without having additional information required to perform a more unambiguous classification (from e.g. field investigations) to be misleading and/or subjective. Instead, in the revised manuscript version, we will offer the following figure in the appendix.




New appendix A: The 12 sites mapped by experts and their locations shown in optical imagery for reference (source: SWISSIMAGE).

13. Figure 4: the cp and icp polygons on the interferograms are indistinguishable.

We improve the visualisation to better highlight the two classes as indicated in this panel:




14. Line 203: Reference the sub-panel related to the folds.

Thanks, we adjust accordingly.

15. Line 344: Fig. 10 (and most other figures) appears much later in the manuscript which means a lot of scrolling back and orth. Please place your figures as close as possible to the subsection where it is first referenced!

Thanks, we will adjust accordingly.

16. Table 3: Is the Hausdorff distance measured after geocoding the identified mass-movements polygons? Are the units then in meters?

The Hausdorf distance is measured still in the range-azimuth system, and hence the unit is still not meters. We will make this more clear by mentioning it in the section 2.3.5:

“The Hausdorff Distance is only defined on the positive-only dataset and in azimuth-range units.”

17. Figure 8:
    (a) Same comment as Figure 4 regarding the visibility of cp and icp delineations.
    (b) You reference Fig. 8 in line 419 and state that the DL segmentation masks show general agreement with the expert mappings. Where are the DL segmentation masks shown?

a) Thanks, we apply the same new design as in Fig. 4.
b) We acknowledge that the text formulation was ambiguous, what we meant was the fringe Likelihood in column 3. The masking that was applied to compare the output to the experts was when the fringe likelihood was >0.5. We will make this more clear in the figure caption as follows: “c) Training labels of coherent phase patterns (cp) and incoherent phase patterns (icp) and the likelihood of detection of the final segmentation model. DL segmentation masks refer to c) when fringe likelihood >= 0.5.”

18. Line 484: "very temporal?"

Thanks for spotting this. We correct to “…and maximise temporal variability to…”

19. Line 496: "external dataset" (singular).

Thanks, we correct accordingly.

20. Line 517: ":augmentation" (lowercase following the double colon).

Thanks, we adjust accordingly.

References:

Agliardi F, Scuderi MM, Fusi N, Collettini C., 2020: Slow-to-fast transition of giant creeping rockslides modulated by undrained loading in basal shear zones. Nat Commun. 2020 Mar 12;11(1):1352. doi: 10.1038/s41467-020-15093-3. PMID: 32165629; PMCID: PMC7067777.

Stark, C.P. and Hovius, N., 2001: The characterization of landslide size distributions. Geoühysical research letters, 2001 March 15th. doi: 10.1029/2000GL008527