Reviewer #1 Comments

We appreciate the reviewer's time and thoughtful feedback. We agree that the readability of the manuscript would benefit from modifications to whether particular items are described in the results, discussion, or conclusion sections. Additionally, we are happy to both expand on the discussion of network science itself, clarifying the unique value this method brings to the problem, and further elaborate on the monitoring implications of our analysis. These revisions will significantly enhance the clarity and overall impact of the work. Below, we respond to each of the comments and questions in detail.

Thank you for your innovative work. I have some comments and questions that I listed below:

Line 2 (and throughout the paper): the term "vulnerability" might be misused in the context of risk science (e.g. no estimate is made of infrastructure fragility). I would argue that the terms to be used are "hazard" and "exposure" depending on the context.

> We had introduced the term 'vulnerability' since we had created a new metric and it seemed appropriate to name it. Given the concerns of both reviewers, we will replace `vulnerability' with `hazard' as the context requires, based on the following definition:
>
> > "Landslide hazard maps indicate the possibility of landslides occurring throughout a given area. An ideal landslide hazard map shows not only the chances that a landslide might form at a particular place, but also the chance that it might travel downslope a given distance." - USGS (https://www.usgs.gov/faqs/what-a-landslide-hazard-map)

Line 13: Palmer (2017) refers to slow-landslide specifically and the definition of which specific process the current paper is addressing is not clear at this point

Line 22: please see my comment above - "mass movement of material (rock, earth, debris) down the hillslope, defined as a landslide event". It would be good to define specifically the process under study here. It is not only a semantic problem as rock fall, landslides, debris flow processes are mechanically and spatially different.

> Regarding both the Line 13 and Line 22 comments: we agree that further clarification of the specific process would be helpful to interpreting the results. Unfortunately, since there are no published studies of these recent landslide

events, except for Paul's Slide, we are not sure if all 4 of these events are associated with slow-moving landslides. To provide the full information available, we will note that all 4 of these landslides occur deeper than 1 m below the surface soil, and clarify that this study encompasses both slow-moving and debris slide processes.

Line 26-28: Can you check the https://blogbigsur (Drabinski and Bertola) reference- I couldn't access the reference mentioned. Any other scientific publication available?

We have confirmed that the website we list https://blogbigsur.wordpress.com/2023/03/31/update-37-with-repairs-underway-travel-opportunites-still-abound-on-the-big-sur-coast/, https://blogbigsur.wordpress.com/2023/03/16/update-31-assessments-on-highway-1-continue-at-areas-damaged-by-most-recent-storms/, and https://blogbigsur.wordpress.com/2023/03/07/highway-1-at-mill-creek-to-open-by-end-of-march-pauls-slide-still-set-for-long-term-closure/ is still live as of May, 2025. These are reports from the California Department of Transportation which provides details of the event. This referenced source provides valuable documentation when formal scientific studies are unavailable. Additionally, the California landslide database found on https://www.conservation.ca.gov/cgs/landslides have recorded these events and we will add that citation to lines 26-28.

Line 29 : More background information about network science would be good to have here - e.g. Nodes, Edges etc. Why is a network framework interesting in this case?

Network science is one of the largest developments in applied math and statistical physics during the past few decades, a field with its own conference (NetSci) which reports on its application to many fields including the granular/amorphous materials (Refs Bassett et al., 2011; Kivela et al., 2014; Mucha et al., 2010; Papadopoulos et al., 2016; Porter & Gleeson, 2016; Porter et al., 2019 in the paper); this large and active community of researchers motivated the present work.

As for a specific motivation for this project, the underlying landslide material is indeed made up of grains, and in our original and successful application (Desai et al 2023) we had been primarily attracted to the ability of network science to reduce a complex problem to a description in terms of spatiotemporal relationships. The network science approach is to provide an overview of the state of the system in terms of a set of nodes connected by edges, where each edge contains quantifiable data about the relationships between the nodes, information which (in the case of landslides) is available from remote sensing data. In our revision to the paper we will emphasize this context (as was done in our earlier paper, but we hadn't originally repeated it here.)

Line 31 : "two categories: stable and vulnerable. A region is considered vulnerable if it is likely to experience a landslide event" - see my comment above - "susceptible" might be more appropriate?

Unfortunately, the word "susceptibility" already has a meaning for researchers studying both landslides and the statistical physics of phase transitions. Therefore we decided not to use that terminology (in fact, this is what led us to the term "vulnerable"). As stated in the earlier comment on the same topic, we will adjust our language usage to use the word "hazard".

Figure 1: a minimap of the location in the state / country would be nice for a geomorphologic context

During revisions, we will add that in as an inset to Fig.1 (a).

Line 64: it is unclear to which "mass downslope movement" you are referring to here. The link between slow moving landslide and evidence of associated mass movement is missing

Given the confusion of 'mass downslope movement' and slow-moving landslides in addition to the uncertainty of the processes of each of the 4 landslide events, we intend to clarify that 'mass downslope movement' encompasses slow-moving landslides (Paul's Slide) and debris or rock slides (potentially Mill's Creek).

Line 70: 1) 40x40 m2 should be 40m2 or 40x40m, I believe 2) The temporal resolution is not explicit

1) We mean (40 x 40) m^2 and will add these parentheses during revision.
2) The temporal resolution is irregular. Sometimes it is 12 days and sometimes 6 days, corresponding to Sentinel-1 passes, and we will add an explicit mention of this during revisions.

Line 90: How much of the total area represents the mask? Can it impact the analysis?

The mask represents about 1.7% of the total area. For any of the 17 sub-regions, the mask represents 4.8% on average, with the min being 0% (sub-region #1) and the max being 10% (sub-regions #7,8,10). We will add a summary of this during revisions. Through tests done on an artificial system to mimic InSAR, we found that this does not impact the analysis since (1) it accounts for a small amount of the area and (2) the network analysis is robust to this choice of mask.

Line 105: A representation of the graph and communities would be important for the comprehension

We have a figure that represents the graph and communities in *Desai et al., 2023* (https://doi.org/10.1103/PhysRevE.108.014901) Fig 3. We will reference this specific figure in the paper.

Line 106: A Poisson sampling with Delaunay triangulation is unlikely to follow a hydro-geomorphologic logic - would there be an advantage in using slope units as the basis for the nodes/edges for example (https://doi.org/10.1016/j.geomorph.2020.107124)?

Thank you for sharing that paper, which is an interesting approach. Our choice of Poisson sampling with Delaunay triangulation achieves a similar purpose of encoding the underlying heterogeneity and objective delineation at a much higher spatial resolution (20 m) than slope units would. By factoring in slope steepness, we incorporate the topography within the weighted network. Therefore, we don't see an advantage, but it would be worth trying in future work.

Line 109: "we calculated the average velocity and slope of any two connected nodes and set that as the edge weight" - Are the community distribution sensitive to a metric different than the average (e.g. Maximum, Skewness, Kurtosis)?

The partition into communities is (by design) sensitive to the metric chosen to set the edge weight, which is why we used physics-based quantities to weight our edges. The average velocity and slope of any two connected nodes captures how the inverse-viscosity and gravitational load influence the system. Through tests done on an artificial system to mimic InSAR, we found that choosing the maximum or minimum would introduce a high-sensitivity to noise inherently presented in InSAR data; this reduced the effectiveness of the method. Additionally, choosing the skewness or kurtosis is more about looking at the variability in these measurements, which is doesn't tell you how much the system is currently flowing

Line 113: More details are needed about the GenLouvain algorithm

The generalized Louvain method (Blondel et al., 2008) is a standard network science technique, with over 16,000 citations. As such, it has become standard not to describe it in detail but instead to refer to the features for which it has been chosen. In the revision, we will add that it was selected because it divides a network into communities by identifying where the edge weights are stronger within the community than one would expect at random, and this is the feature we are looking for within our data .

Line 125: In the case of a catastrophic failure (Millions m3), I would expect several communities involved to remain stable while bordering communities would see an increase in Z values - correct?

> No, what we observe is that bordering communities see a decrease in Z-value between time slices while involved communities see an increase in Z. Communities involved with the moving hillslope represent areas that have increasing velocity. This translates to a higher than average weight within the network (weight is velocity times slope). Since communities are partitioned by the mean weight of the network, an area that is moving faster than average is identified as a community, and will continue to be identified as long as the area has a higher velocity. The Z-value for that community will increase then. Areas that are moving slower than the mean will not be identified as a community, and will therefore see a decrease in Z-value.

Is there any (albeit rare) scenarios where the Z-score could, for example, average out and be misleading?

> That is an interesting question, but we struggle to come up with a scenario in which this would be possible for the following reasons. The Z-score is computed for the entire layer instead of for each community in the layer, and it only measures the positive persistence of any community. Since the Z-score does not account for communities decreasing in Z-value within a layer and we look at the overall Z-score change between layers, we are not sure if there is a scenario which could be misleading.

Line 138: Can you be more explicit of what a Z < 0 would actually mean?

> Z_t considers the change in Z between any two time slices. So if Z decreased in the following time slice, then $Z_t < 0$, but if Z increased, then $Z_t > 0$. When Z_t is negative, there is very little persistence in communities in time. In this system, this corresponds to the dry season where there is very little forcing detecting in the hillslope. See our response for line 125 for a more in-depth explanation. We will add this description to the paper.

Line 140-145: It is not clear to me over which period the average community persistence is calculated; is it a rolling average?

No, it is not a rolling average, but an average taken over the entire time. We will clarify that the average community persistence is taken over the entire period of 2015-2022. This could, of course, also be done as a rolling average if the technique were deployed as a monitoring tool.

Line 148: "Here, darker sub-regions represent higher peak Z-scores. sub-regions with a relatively stable Z-score had peak Z < 2.5. Within the sub-regions that showed increasing Z, some sub-regions have peak Z < 3, and some have peak Z " - can you be more explicit about the 1 to 4 categories shown on Figure 2c?

The 1 to 4 values are not categories, but rather a binning of the calculated peak Z score for that sub-region. We will modify this line of text to clarify that these are numerical ranges of values, and will connect the peak Z values defined in Fig. 2c to the results more clearly.

Figure 2b: the asterisks are really small

We will increase the size of the asterisks.

Line 159: The Multivariate analysis should probably be introduced in the methodological section with the result explained in the Result section. A Discussion section could then be added before the conclusion

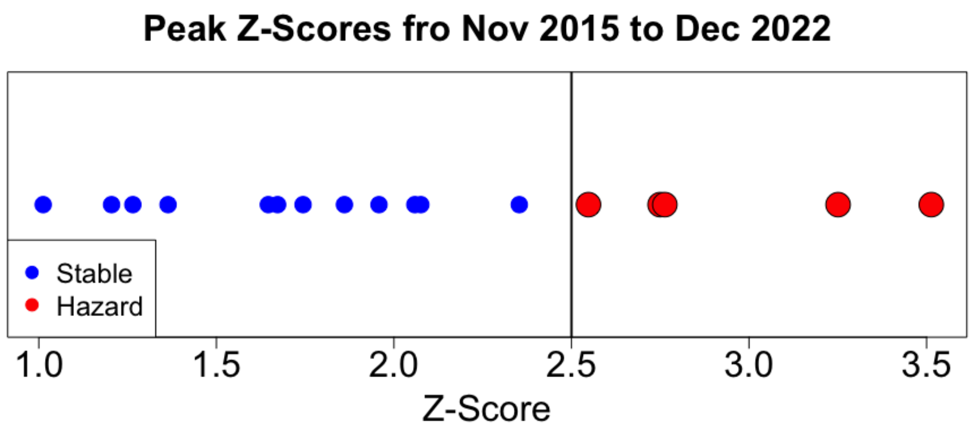
We believe that introducing the reasoning and method of the multivariate analysis in the methods section makes sense. We will move the first paragraph in section 4.1 to the methods section with some rewording.

4.1 Multivariate Analysis: Can you clean this paragraph, as there are several discrepancies and it makes it hard to follow: e.g. "Community persistence exhibits positive correlations with mean displacement (-0.53)", "Moreover, precipitation has a strong positive correlation by 0.63 with precipitation"

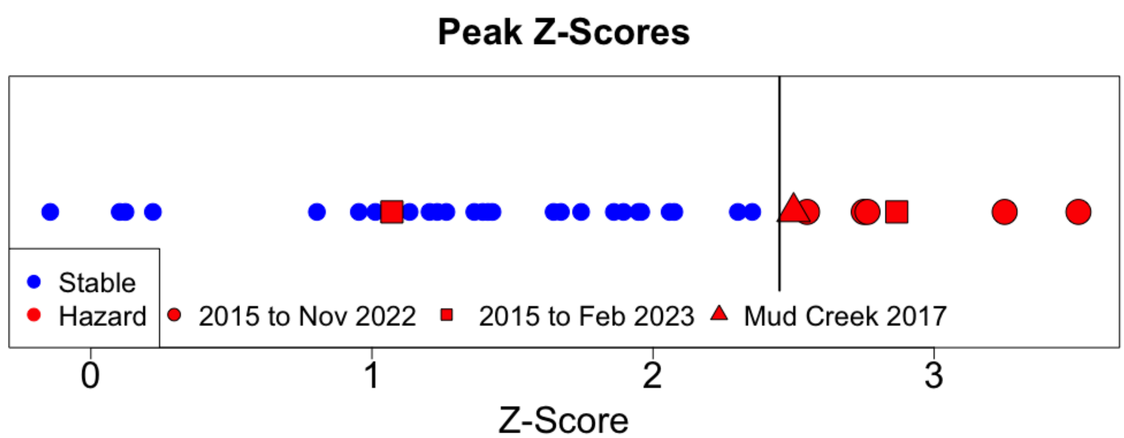
We will revise this paragraph to make it clearer. We mean to state the precipitation has a strong positive correlation of 0.63 with the landslide events.

Figure 3: Could you add a ROC plot and AUC score from the Z-score and events?

Below, we plotted the peak Z-score for each of the 17 subregions, calculated for the Nov 2015 to Dec 2022 dataset which is the focus of this paper. Red indicates a landslide event, and blue indicates a stable region. There is a sharp boundary between high-Z (Z > 2.5) and low-Z (Z < 2.5) events, such that an ROC plot would simply be a step function. We will include this information in the revised manuscript or Supplemental Material, either as a plot like this or in some other format.



If we instead combine the analyses from both the Nov 2015 to Dec 2022 and Nov 2015 to Feb 2023 periods, as well as separately plotting the Mud Creek event from 2017, then a similar plot [see below] shows only one Z-score which differs from that classification. This also makes an ROC plot an inappropriate tool: for any reasonable choice of threshold near Z=2.5 there is just this one outlier.



The outlier is the Gilbert's Slide, which was identified as a hazard in the Dec 2022 period, but fell below the threshold a month before it catastrophically failed (mentioned on line 157). All other events fall above Z = 2.5.

Line 202: "we analyzed two time periods:Nov 2015 to Nov 2022 and Nov 2015 to Feb 2023" the two analysis periods remains unclear in term of their relationship or purpose

Yes, the two time periods were chosen to show how analysis before and after the landslide events compare. We will add clarifying language to the paper to make that clear.

Line 206: The 97% is coming out of the blue and not convincing (as you pointed out) and consider, from what I understood, a single threshold. See above my comment on the ROC curve. Could you iterate the threshold with various scoring metrics to identify an optimal threshold and provide a better selling point for your method?

The 97% comes from the calculation that during the two time periods we ran this analysis for: Nov 15 to Dec 22 and Nov 15 to Feb 23, there was only one time when the analysis mis-identified a region as stable: Gilbert's Slide. This region was correctly identified as a hazard in Dec 2022, but fell below the threshold in Feb 2023 right before it catastrophically failed in Mar 2023 (but only when using the longer time period)

For both of the (overlapping) time periods, each of the sub-regions is either classified as stable or a hazard, depending on if a region experienced a landslide event in the months following the analysis. The 97% comes from 1 false negative out of 34 measurements (with a statistical complication that we are including each sub-region twice, once for each fresh analysis).

The reason for the single threshold is described above, and there is not another choice that is reasonable to make and would adjust the observed percentage. Since there is only one subregion (out of 17) causing the false negative, and its Z-score is far below any reasonable choice of threshold, considering sensitivity and specificity would not improve our classification: this subregion is an outlier.

We will make this calculation, and the lack of flexibility to choose a better threshold, clear during revisions.

Line 217: Correct "(Oregon State Univeristy, 2015) with velocity and slope in S2" - presumably Supplementary Information 2?

Yes. We will correct that.

Line 221: What is the result of the WRF-Hydro model doing in the Conclusion section? it seems out of place

This placement arose because we observed a negative result, and this placement originally seemed like a way to mention it more briefly than we did for the positive results. Since both reviewers have suggested that it be included as a main result, we are happy to instead include it in the Results section.

The Conclusion looks more like a Discussion + Conclusion and the conclusion lacks specific recommendations for practical implementation. Overall the landslide inventories and their uses remains unclear and need to be addressed

In the revised manuscript, we will reorganize these sections to distinctly separate the discussion of the results from the conclusion. Additionally, we will expand the Discussion section to more thoroughly interpret the results and contextualize the methods used, especially incorporating the specific discussion points raised by both reviewers. We will place emphasis on how the network science techniques simplifies a complex system and adds to identifying the transition from stable to hazardous, as well as add in more about the potential applications of this method as a monitoring tool. We appreciate the thoroughness of your comments to clean up the language for better clarity throughout the paper.