**Summary and recommendation**

This work presents a simple yet novel method for characterizing landslide potential at the regional scale using a normalized rainfall index, A*. The concept is based in part on the idea of a characteristic recurrence interval for landslide inducing storms proposed for global application by Marc et al., (2019, 2022), but introduces a substantial advance by considering the relative soil-saturation rather than a rainfall recurrence interval using the Antecedent Wetness Index (AWI), implemented for Seattle by Godt et al. (2006) and similar to a leaky-bucket approach used for the SF Bay Area by Wilson and Weiczorek (1995). The authors demonstrate the utility of this parsimonious approach across the entire state of California, a region of the U.S, with considerable landslide hazards and risk, as well as extensive event-based inventories and rainfall data to apply and test their model. The work demonstrates a nice correlation between their index exceedance and landslide occurrence as predicted by their Landslide Potential Area (LPA) indicator based on a calibration with four events and a total of nine events for demonstrating the validity of the approach.

The work is highly relevant and will be of substantial interest to landslides researchers as well as potential applications for situational awareness and near-term planning in the decision-management arena. In particular, it presents a nice contribution in terms of moving from rainfall-only driven triggering to consider the role of the subsurface and antecedent conditions at the regional scale, albeit through a simplified bucket model. This is in fact remarkably well-aligned with the type of approaches we propose in a recent commentary that is currently open for discussion (Mirus et al, NHESS-Discussion, 2024). I was pleased to see that the authors demonstrated that their non-dimensional index A* appears to be more valuable than the IVT criteria of ARs, which is often called upon as a proxy for landslide potential. So, ultimately, this is a great proof-of-concept for going beyond storm metrics used in the atmospheric sciences and adding more landslide-relevant considerations. It will surely be a nice contribution as an approach that is likely to gain traction moving forward.

Overall, the manuscript is very well-written, and the figures and tables are clear. The motivation and background are well-supported, and the work builds upon other novel contributions in the literature dating back to early contributions from previous USGS researchers (i.e., Campbell, 1975; Wilson and Wiezcorek, 1995; Godt et al., 2006) among others. The paper is short, which is nice for readability, but as a result I found there are some critical details that are missing from the methodology and some other areas for discussion that would help readers better appreciate and repeat the work. These related to both the advances and current limits of the assumptions baked into the AWI approach as implemented as well as those presented by data availability that limits more rigorous testing. These gaps can likely be addressed through the major revisions outlined in the attached text, including line-by-line comments. A revised manuscript that addresses these points should ultimately be accepted for final publication in NHESS.

**Soil moisture and AWI as a proxy**

- It is interesting to use a soil-moisture threshold, when previous work at least for many of the SF Bay Area storms, suggests that positive porewater pressures are a more relevant metric than soil moisture levels for actual triggering since near-fully saturated soils are needed to trigger landslides (e.g., Thomas et al., *Landslides*, 2017). Furthermore, the AWI concept as presented in Godt et al. (2006) was used to modulate the Intensity Duration (ID) threshold, not as a stand-alone predictor of landslides. This is important to note, particularly given the conclusions that high-intensity rainfall and other mesoscale processes may be relevant for triggering and are not consistently captured in the A* metric. Potentially this approach could be improved even further by combining A* with short-term forecasts of precipitation intensity. For example, the AWI is now used with Quantitative Precipitation Forecasts of intensity and duration threshold to provide situational awareness for the greater Seattle area (see https://www.usgs.gov/media/images/seattle-washington-landslide-monitoring-site-awi).
- As the authors note, the AWI is designed to fill quickly since rain is added instantaneously, and it also drains more quickly than real soils (see below). So, in many ways it is more appropriate as an antecedent metric and not as a triggering metric. In this context, why would you expect this to work better than rainfall-only thresholds identified with other modeling approaches (e.g., Scheevel et al., *USGS*, 2017; Patton et al, *NHESS*, 2023)? We have some discussion about this in a paper about Seattle (Mirus et al. *Landslides*, 2018) and something similar in scope might be useful to flush this topic out a little more fully in the discussion.
- Most importantly, it appears spatial variability in the AWI calculations is entirely a result of variable rainfall across the State. Did I miss something? If not, this is a very significant simplifying assumption that needs to be stated more explicitly. I understand that the point of this paper is largely as a proof of concept for the simple index approach, and it does that successfully, but it's not clear how the technique could / should be implemented for a given site of interest. Here are a few additional points to consider:

    1. Technically the A* is not a model parameter, it's an index/proxy simulated by the model. However, the AWI does use at least one parameter, kd, which wasn't discussed. It is a lumped fitting parameter that reflects the soil drainage rate, which integrates many the soil properties (Ksat, retention curves, porosity, thickness, etc.), as well as other geometric factors that affect drainage (slope, convergence, soil thickness, etc.). Although the AWI is indented to be representative over a broad area and assumes infiltration is added instantaneously, there is no reason to assume that <u>all the soils in California</u> will drain at the same rate. In the revisions, please discuss why this assumption is ok enough for your objectives and how it potentially influences the results.
    2. The equation in Godt et al. (2006) accounts for ET in calculating effective rainfall Ii. Did this study also use ET? If not clarify that and explain why. Godt et al. (2006) used a fairly simple monthly estimate of PET, which was good enough for calculating antecedent conditions at the hourly scale, but has a major impact on the outcome if ignored, particularly during spring months when ET ramps up.

3. Given that the USGS has published soil moisture data for at least one of the BALT sites with some landslide-inducing storms that you considered (i.e., Thomas et al., USGS-Data, 2017), have you looked at how these compare qualitatively to your AWI calculations for those grid cells? Godt et al. (2006) developed their parameterization of the AWI based on observed soil moisture data from Edmunds, WA (refer to Figure 4 in their paper). With sufficient calibration, one can get an AWI that nicely matches observed soil moisture dynamics for a region of interest (for example, see below).

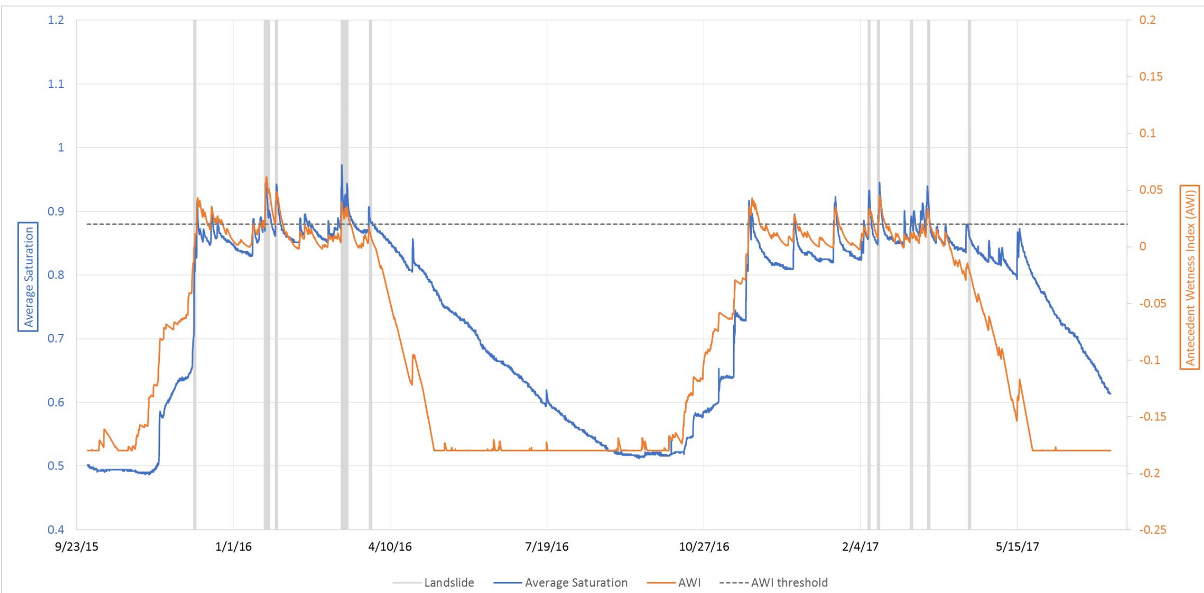


**Figure**: time series plot of AWI calculated using parameterization from Godt et al. (2006) as displayed on the USGS website ([https://www.usgs.gov/media/images/seattle-washington-landslide-monitoring-site-awi](https://www.usgs.gov/media/images/seattle-washington-landslide-monitoring-site-awi)) versus mean observed saturation from in-situ soil moisture sensors in Mukilteo, WA (refer to Mirus et al., *WRR*, 2017). The AWI wets up more quickly at the beginning of the rainy season, then matches observations for the higher saturated values reasonably well, but not as well during the dry season as the soil drains below field capacity and approaches wilting point. Whereas at the event scale the AWI drains too slowly relative to real soils, at the seasonal scale it drains too quickly.

**Data limitations and false/failed alarms**

- The period of record with only 15-years does not adequately support the conclusion of a 15-year recurrence interval for identifying A* and suggests a longer record is needed to better constrain this recurrence-interval based ratio. I didn't quite understand why the full 19 years of the rainfall dataset wasn't leveraged fully. Additionally, although the input rainfall data is time limited, one could easily calculate the AWI going further back with different time-series of rainfall databases to get a sense of whether 15 years is close enough or if much longer records are needed to capture this relevant recurrence interval.

- Any criteria used for landslide assessments, whether temporal, spatial, or spatiotemporal should in some way consider the potential for false or failed alarms. I understand that there are data limitations that likely preclude a rigorous quantitative analysis of performance or even explicit consideration of failed alarms or false alarms, but some discussion is needed. Can you identify at least a few "big" storms that might have triggered landslides and did not? Were they also associated with positive (or high) LPA values? Alternatively, from the storms you did evaluate are there any substantial areas with some LPA that *didn't* produce any landslides (probably, given your very conservative 5-degree slope threshold), and if so, how much?

- Can you correlate the LPA to landslide densities, numbers of landslides, or even just a general qualitative descriptor of how widespread landslides were within the areas predicted? Does a higher LPA link to a greater number/severity of landsliding? I ask this because the magnitude varies so widely and the slope is so low, that it seems some guidance would be needed to understand what the difference might be between LPA = 60 vs. LPA = 11k. Given the conservative slope threshold, I could easily imagine a scenario in which lower LPA in a localized area of very steep terrain triggers a higher density of landslides than a broader area of less steep terrain with a much higher total LPA.

**Specific line-by-line comments:**

12 – Can you provide a few words that better characterize A* for the abstract? Is it a proxy, or a dimensionless index that you test as a proxy for landslide probability? It's not clear until I read the paper.

54 – Consider mentioning however, that most of those ARs from their analysis did NOT result in landslide triggering. So, there's more going on than just high IVT and ARs.

87 – I'm not sure that this Bogaard and Greco paper is an appropriate citation for the ID threshold since it largely focuses on the inconstancies and advocates for moving away from the approach.

115 – It's worth citing some of the other recent literature using direct measurements of soil water content to identify empirical thresholds (see Mirus et al., NHESS Discussion, 2024 and references therein).

156 – Technically this isn't a parameter, but rather an index or nondimensional ratio that is the output of your calibrated/parameterized AWI model.

178 – Godt et al. (2006) used an effective rainfall minus PET to calculate Ii, so that's an important consideration especially for California, particularly to consider potential impacts of climate change.

195 – Consider the Thomas et al., (*WRR,* 2019) not only showed that SMAP over-smoothed the soil moisture data, but PRISM also over-smoothed out the orographic effect on precipitation intensity and totals. Your dimensionless approach with A* should help avoid this issue of magnitude of precipitation (or soil moisture) measurements, which is a common issue with satellite precipitation estimates.

203 – Again, 15-years seems like a short window considering your conclusion and that Marc et al. (2022) found that the 10-year rainfall anomaly was potentially only 50% effective in characterizing landslide potential for extreme and widespread landslide events. Some discussion or further analysis using a proxy with older rainfall records for one (or more) location seems prudent to better support your conclusion.

207 – Technically you're not calibrating A* since there isn't any parameter adjustment, you're identifying which value of A* corresponds to landslide occurrence in the same way one might use a critical factor of safety for estimating slope failures.

Also, refer to general comment about kd. Was the soil variability considered in the field capacity, etc. considered for calibrating kd in area without monitoring data or landslide occurrence data?

223 – Consider specifying how the interpolation is applied (linear? Krigging? Other?) rather than citing a toolbox and assuming everyone knows how it works.

225 – Why not use a 19-year time series to identify the anomaly?

236 – Why not use a more sophisticated approach than 5-degree slope? For example, the Godt et al., *NASL*, 2012) model is available at 1km resolution and Brabb et al, *USGS,* 1999 found a 25-degree slope was suitable for rapid shallow landslides and debris flows. Within this context your threshold is very conservative. Discuss?

Figure 2. This relationship does not appear to be 1:1 and suggests that your AWI might need to vary for different landslide triggering storms or locations (see general comment #2 about calibrating kd). How do your other 5 storms plot here?

423 – Do you mean Luna and Korup, 2022?

247 – What about false alarms?

266-67 – I don't quite follow this. Isn't the maximum AR value used going to be independent of location to produce the value and time of AR max? How does it correspond to the location of one or more observed landslides?

Table 1 – This includes results and should probably be linked to those later in the paper. Also, can you list the number or density of landslides you showed in Figure 8c (albeit, maybe acknowledging that the inventories are incomplete)?

What about non-events? Are you able to introduce any of these, even anecdotally?

Figure 4 – This is a European/international journal, so I suggest you clarify that the numbers are years (e.g., January 2005, not January $5^{th}$)

453 – Again, a good place to point out that Cordeira found that while most landslides were associated with ARs, most ARs didn't necessarily produce landslides.

465 – Maybe I missed it, but your dataset considered 6h rainfall, so how does that account for the high-intensity bursts?

Fig 8c. This is the most compelling argument for the broad utility of this approach. You might be able to do even better if your kd values varied by soil type and ET were considered in calculation of AWI.

540 – Here or elsewhere warrants some discussion of false positives and false negatives, which are not tested thoroughly. Even if this is justified by a lack of comprehensive data, this is still a limitation of the approach worth discussing (if only to point out the data needs for further developing this method!).

548 – Again, what about local variation in soil properties (depth, ksat, retention curve, etc.) all of which are wrapped up in the kd parameter that you seem to have held constant for the entire State of California?!

569 – This potential value underscores the need to consider ET in the effective rainfall Ii.