

We are grateful to the editors and all three reviewers for taking the time to provide feedback on our manuscript. The comments and suggestions will greatly improve the overall quality, clarity, and rigor of the paper. We have responded to all reviewer comments and plan to add all revisions to our manuscript. Following the instructions from the editor, we have not made all revisions yet, but we will ensure that all revisions are clearly reflected in the updated text. Below, the original comments appear in bold roman typesetting, and our responses are provided in italic typesetting.

Comments from reviewer 2

This study introduces the concept of incorporating normalized soil moisture (NSM) into rainfall-triggered landslide threshold models to enhance early warning performance. The integration of long-term inclinometer data, NOAA precipitation records, and NASA SMAP soil moisture observations provide a valuable multidisciplinary framework. However, from the standpoint of scientific rigor and generalizability, the manuscript still presents notable limitations in terms of data representativeness, statistical validation, spatial-scale consistency, and geotechnical mechanism interpretation. Major revision is required before the manuscript can be considered for publication. The authors are encouraged to strengthen the temporal resolution, address spatial resolution discrepancies, and provide quantitative uncertainty analyses to improve the reliability and applicability of the conclusions.

We thank the review for their thorough review of our manuscript. In the revised manuscript, we improve the discussion of the limitations to better qualify our conclusions and avoid overstating the applicability of our findings. We will also clarify that we did not intend to present generalizable relationships, but rather a transferable framework that can be applied to other regions where sufficient data is available. We recognize that our data has limitations both temporally and spatially, but we are unable to change these. We do plan to more clearly acknowledge these limitations in the revised manuscript.

We intend to address each of the specific comments that reviewer #2 has identified through revisions to the manuscript. Our responses are detailed below.

1. **The manuscript frequently uses the first-person pronoun, which should be removed for formal scientific writing. For instance, “We aim to improve existing rainfall thresholds for landslides along highways by incorporating antecedent soil moisture conditions.” should be revised to “This study aims to improve existing rainfall thresholds for landslides along highways by incorporating antecedent soil moisture conditions.”**

Response to Comment 1: We thank the reviewer for this suggestion. It is our understanding that both first-person and third-person constructions are acceptable within the journal, and we have tried to maintain a consistent style with other literature on this topic. We believe that this is more a matter of preference than a technical comment and so have respectfully decided to maintain our current style. We will, however, review the final manuscript carefully to ensure that our writing style is consistent and does not detract from the technical contributions of the manuscript.

2. The use of quarterly inclinometer readings results in low temporal resolution.

This makes it impossible to accurately link specific rainfall events with landslide occurrences, severely weakening the temporal correspondence between rainfall-duration (I-D) thresholds and soil moisture response, and consequently reducing statistical correlation and causal interpretability.

Response to Comment 2: Thank you for this important comment. We acknowledge that the use of quarterly inclinometer readings results in low temporal resolution. This is, unfortunately, an inherent limitation of many landslide databases, where the exact timing of slope movement is rarely known with high precision. We agree that this weakens the statistical correlations between landslide events and specific storms and will clarify this limitation in the revised manuscript.

3. The manuscript does not clarify the training background or consistency evaluation of the operators involved in inclinometer data acquisition, which could introduce subjective bias.

Response to Comment 3: We agree that this information is important. In the revised manuscript, we will clarify that all inclinometer readings were collected using the same inclinometer probe by the same trained geologist who has more than 10 years of experience in landslide monitoring in Alabama.

4. The spatial resolution of the SMAP data (9 km × 9 km) is too coarse to represent local-scale topographic and soil moisture variability, limiting the applicability of the model in mountainous areas with small-scale landslides.

Response to Comment 4: Thank you for this comment. We agree that the 9 × 9 km spatial resolution of the SMAP dataset is relatively coarse and may not capture local-scale variations in topography or soil moisture. We selected products with coarser spatial resolution because they provide the temporal resolution needed to analyze and predict landslide behavior. We will add text to better explain this rationale at line 191. Nonetheless, SMAP was selected because it provides publicly available, continuous, spatially consistent, and long-term observations that are appropriate for regional-scale assessments where finer-

scale moisture measurements are unavailable. The objective of this study was to develop a landslide prediction threshold that could be applied across Alabama. While collecting site-specific, ground-based soil moisture data may provide a more accurate threshold, it would severely limit the broader application of the threshold to locations where *in situ* soil moisture data is collected. In this study, the SMAP root-zone soil moisture product is intentionally used as a regional indicator of average wetness, not as a site-specific measurement or a proxy for matric suction at the depth of the shear zone. The manuscript explicitly states this distinction (Line 193-195). The use of normalized soil moisture further reduces bias associated with local variability by expressing moisture conditions relative to each grid cell's long-term average, making comparisons across the state more robust. We will clarify these points in the revised manuscript, and we will also note that future work could incorporate higher-resolution soil moisture datasets or site-specific field measurements to improve representation of local-scale hydrologic variability.

5. The sample size and regional representativeness are limited. The dataset is small and highly localized, resulting in weak generalizability of the conclusions and making it difficult to substantiate the claim of a “transferable threshold.”

Response to Comment 5: We agree that the term “transferable threshold” was not appropriate given the regional nature of our study. It was not our intention to create a generalizable threshold, but rather to present an integrated approach to develop thresholds when sufficient datasets are available. In the revised manuscript, we will instead use the phrase “transferable methodology” rather than “transferable threshold,” as this more accurately reflects the intended contribution of the study. The approach we present can be applied elsewhere, but the specific threshold values should indeed be considered region-specific, and this will be clarified.

6. Figure 3 fails to clearly demonstrate the relationship between rainfall, soil moisture, and landslide occurrence. The graphical evidence does not convincingly support the authors' interpretation and should be clarified with enhanced visualization or statistical quantification.

Response to Comment 6: Our intention with these figures was to provide motivation and context for our approach and not to demonstrate statistical correlation between these values. We will revise the text to better describe the intention of these figures.

7. Geological and soil parameters were not quantitatively controlled. Although the manuscript classifies strata into three lithologic types, key physical parameters (e.g., permeability, cohesion, plasticity index) are not incorporated, which weakens the geotechnical basis of the proposed thresholds.

Response to Comment 7: We agree with the reviewer that our approach does not control for differences in geotechnical properties between the different sites. The purpose of the lithological classifications was to provide context for the types of geologic units that are susceptible to landslides in Alabama. We agree that developing site-specific geotechnical properties is important for analyzing single slope failures or designing repairs, but our aim is to provide a regional assessment of timing of movements at already unstable locations. We would see our thresholds as a compliment to geotechnical slope stability analyses and not a replacement. We will ensure this motivation is clearly described in the introduction of the revised manuscript.

8. The innovation is more phenomenological than mechanistic. The study focuses on statistical correlation without sufficient exploration of the underlying hydro-geotechnical processes that govern the observed trends, reducing the theoretical depth of innovation.

Response to Comment 8: We agree that the present study is primarily phenomenological in nature, as our focus is on developing empirical relationships between rainfall, soil moisture, and slope movements across a relatively large region. We believe this is consistent with the aims of the journal, which include regional analyses that document observable patterns, improve understanding in data-limited environments, and provide a transparent, citable foundation for future scientific development. Our study provides an initial, data-driven assessment that documents observable regional patterns in Alabama's highway slopes. The intention is not to replace mechanistic hydro-geotechnical models, but rather to establish an approach to provide regional-scale assessments of when instabilities are likely to occur. We will clarify this motivation in the introduction to our manuscript.

9. The applicability boundaries of the proposed approach are not explicitly discussed. This omission reduces the methodological rigor and limits understanding of the model's valid domain.

Response to Comment 9: Thank you for this comment. In the revised version, we will clarify the valid domain of the methodology, emphasizing that the developed thresholds are specific to monitored highway slopes in Alabama and reflect the region's geological, hydrologic, and climatic conditions. As noted in our response to Comment 5, the threshold values themselves are not intended to be transferable; rather, the contribution of the study lies in the transferable methodology used to integrate rainfall, soil moisture, and displacement records showing the empirical workflow can be applied in other regions.

10. The rainfall event classification is overly simplified, and the threshold selection may be too lenient, as it does not account for the independence of consecutive dry periods or short-duration, high-intensity rainfall events.

Response to Comment 10: We agree that the previous discussion of event classification was unclear. In this study, rainfall events were separated following the approach of Leonarduzzi et al. (2017), which defines an independent rainfall event as a sequence of consecutive days with more than 1 mm/d of precipitation. This 1 mm/d threshold ensures that annual rainfall totals are not unrealistically reduced and that multiday storm durations remain physically meaningful. Under this definition, a dry period of at least 24 hours (i.e., a day with < 1 mm of rainfall) indicates the end of one event and the beginning of another, meaning that consecutive storms separated by a full dry day are treated as separate events. Using this event-separation algorithm, all rainy days without a dry period in between were grouped into the same storm. For each identified storm, cumulative rainfall was computed by summing all daily precipitation within the event, the number of rainy days was taken as the event duration, and rainfall intensity was calculated as cumulative rainfall divided by event duration. This provided a consistent metric for rainfall intensity across all events. We acknowledge that this daily resolution method does not explicitly account for very short, high-intensity bursts or the independence of shorter dry periods, as more advanced approaches require sub-daily rainfall data. However, the availability of hourly meteorological data is limited to a small number of sites and time periods, so a threshold based on sub-daily rainfall data would have limited use as a predictive tool. We will ensure that the revised manuscript more clearly explains this approach and acknowledges the limitations inherent in using daily rainfall.

11. The discussion section focuses primarily on whether prediction accuracy improved, but lacks an in-depth analysis of data sources, model structure, and scale compatibility. A more comprehensive discussion of these factors is needed; substantial revision of this section is recommended.

Response to Comment 11: We agree that the discussion section should more explicitly address the limitations in the study identified in both this and other comments. In the revised manuscript, we will substantially expand the *Discussion* and *Limitations* sections to incorporate these elements.

Specifically, we will add:

- **Data source limitations**, including the spatial resolution of SMAP (9 × 9 km), the daily resolution of precipitation records, and the quarterly temporal resolution of inclinometer measurements.

- **Model structure limitations**, noting that the thresholds are derived from empirical relationships and do not explicitly consider the hydro-geomechanical processes or site-specific material properties (e.g., permeability, cohesion, shear strength) governing slope stability.
- **Scale-compatibility considerations**, clarifying how regional-scale hydrologic indicators (rainfall, SMAP soil moisture) interact with highly localized slope movements measured by inclinometers, the tradeoffs between temporal and spatial resolution in gridded data products, and why this mismatch affects interpretability and transferability.

12. Performance metrics are reported without confidence intervals or statistical significance testing. Consequently, the claimed improvements cannot be validated statistically. The authors should incorporate cross-validation or independent testing.

Response to Comment 12: Thank you for this suggestion. In response to this comment and Comment 4 from Reviewer 3, we computed receiver operating characteristic (ROC) curves and AUC values for all five thresholds (Figure R1). In addition, we evaluated the thresholds using the Matthews Correlation Coefficient (MCC) to assess classification performance under a fixed operational decision rule (Table R1). The results show that both NSM-dependent thresholds (A and B) achieve higher AUC values than the three existing thresholds, indicating improved overall discrimination ability. When evaluated using MCC, Thresholds A and B also outperform the Godt et al. model, reflecting a more favorable balance between true and false classifications by incorporating normalized soil moisture. While the AUC differences among the highest-performing thresholds are modest, the combined use of AUC and MCC demonstrates that the NSM-dependent thresholds provide both strong discrimination and improved operational performance. This combined evaluation framework will be reflected in the revised manuscript.

Table R1. Comparison of landslide threshold performance using area under the ROC curve (AUC), Matthews Correlation Coefficient (MCC), and confusion matrix components.

Method	AUC	MCC	TP	FP	TN	FN
Threshold A	0.983	0.824	100	27	336	5
Threshold B	0.9852	0.805	103	36	327	2
Godt et al. (2006)	0.98	0.765	102	44	319	3
Marino et al. (2020)	0.980	0.843	94	15	348	11
Guzzetti et al. (2008)	0.968	0.549	104	121	242	1

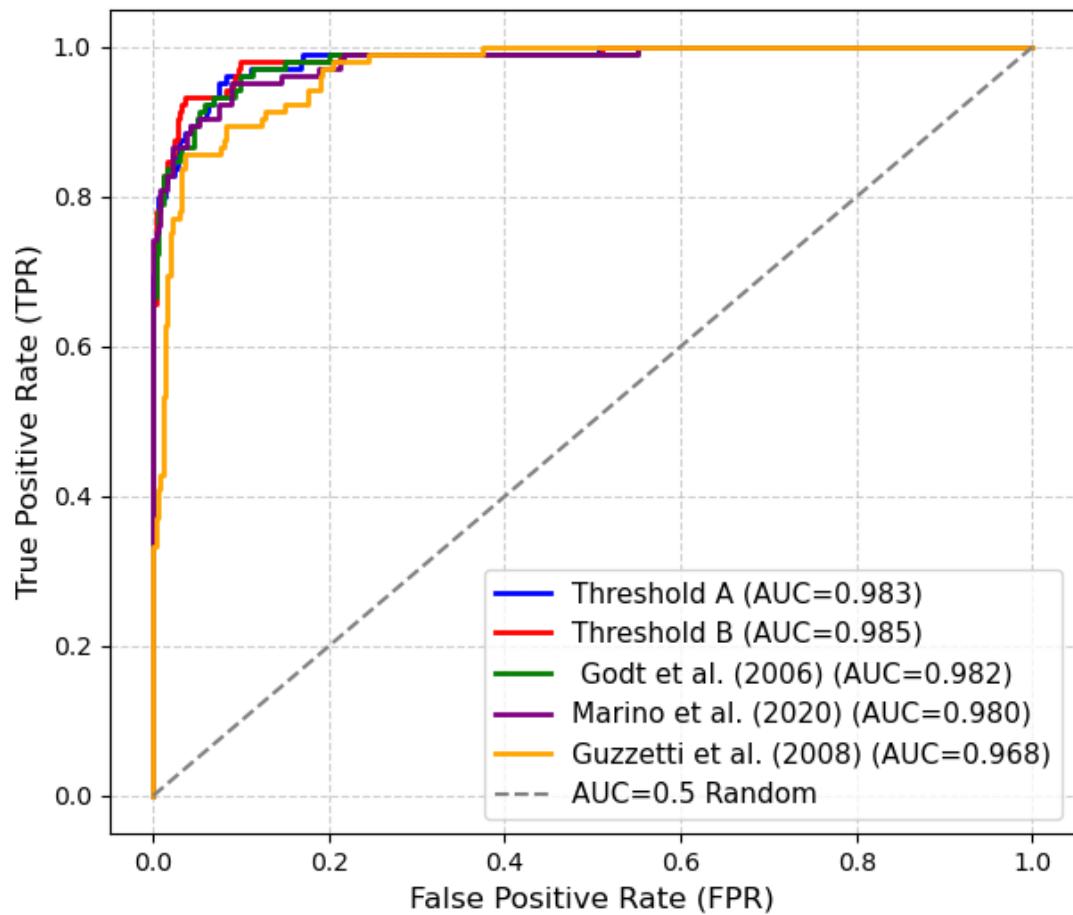


Figure R1. Receiver operating characteristic (ROC) comparison of five intensity-duration thresholds: the two NSM-dependent thresholds (A and B), the Godt et al. (2006)

threshold ($\alpha = 1$), the Marino et al. (2020) threshold, and the Guzzetti et al. (2008) threshold.

13. The normalization procedure may obscure extreme moisture conditions. Averaging across long periods can reduce contrast between very wet and very dry states, thereby weakening the detection of extreme antecedent conditions that critically influence landslide initiation.

Response to Comment 13: We agree that our previous explanation of the normalization procedure was unclear. For applying the NSM-dependent thresholds, we did not perform any averaging but rather used the soil moisture value from the start of the storm event. We did normalize the value using the long-term average, but the same average was used to normalize all of the readings for a given location and therefore would not obscure extreme values. We will clarify this in the revised manuscript.