Response to Referee #1 (RC1)

We thank Referee #1 for a careful and critical assessment of our manuscript. We acknowledge the concerns regarding the validation, operational applicability, and interpretation of the results. We agree that the present framework should be more appropriately described as a prototype with potential for landslide early-warning applications rather than a fully validated operational early-warning system. We will carefully revise the manuscript to address these concerns and provide a more balanced interpretation of the findings.

> **General comment.** *This is an interesting and potentially useful framework, but the manuscript currently overstates its performance. In its present form, ALERT is a promising web application, not a validated operational landslide early-warning system.*
>
> **Comment 1.1.**
>
> *The validation does not appear independent. The same 104 dated landslides seem to be used both to derive the rainfall threshold and evaluate it.*

Response: We thank the reviewer for highlighting this important concern. To strengthen validation and ensure independence from the threshold development process, we plan to incorporate an additional dataset of 50 landslide events not used in threshold calibration. We will clearly distinguish the calibration and validation datasets and document their temporal independence. The revised manuscript will include a detailed description of the validation procedure and the resulting performance. We will also revise the texts so that the claims match the evidence.

> **Comment 1.2.** *The main results are based on only 10 unique landslide days and 10 control days. Sensitivity of 0.80 therefore means only 8 of 10 days were detected. This sample is too small to support strong claims of reliability.*

Response: We acknowledge the reviewer's concern regarding the limited number of unique landslide and non-landslide days considered in the evaluation. We agree that the current sample size is insufficient to support strong conclusions about the framework's reliability. We will expand the validation dataset, where suitable observations are available, and explicitly report the number of independent landslide and control days. We will also discuss the uncertainty associated with the performance metrics and moderate the conclusions accordingly.

Planned changes in the revised manuscript:

- The number of landslide days will be increased through the enlarged inventory.
- Control days are not limited by inventory availability. Instead of 10, we will use all monsoon (June–September) days at each landslide location with no reported landslide within ±3 days. This will give stable estimates of specificity and false-alarm rate, and we will also report precision at a realistic event base rate.
- Results will be presented as an initial regional evaluation of event-day discrimination, and the conclusions will be moderated accordingly.

> **Comment 1.3.** *The ECMWF forecasting component is not actually validated. Figure 10 is only a visual comparison, without quantitative forecast metrics or evaluation at different lead times.*

Response: We agree that Figure 10 is a qualitative illustration and does not validate the forecast component. We will add a quantitative evaluation of both the forecast rainfall and the warnings derived from it. We will strengthen this section by incorporating quantitative performance comparisons between the IMD dataset, IMERG and ECMWF and will report performance metrics, including rainfall bias, mean absolute error, and root mean square error, where suitable reference observations are available. We will also examine forecast performance at different lead times, subject to the availability of corresponding forecast and observed rainfall data. Limitations will be explicitly acknowledged in the revised manuscript.

- Figure 10 will be kept but explicitly labelled as illustrative. The statement in the conclusions that bias-corrected forecasts “provide reliable rainfall forcing” will be replaced by the quantitative results.

Comment 1.4. *The threshold methodology is unclear. The manuscript gives conflicting descriptions of the regression, and assigning unknown landslide times using the maximum rainfall during the day may introduce serious bias.*

Response: We agree on both points. We will revise the methodology section to provide the regression equations and describe the fitting procedure.

Planned changes in the revised manuscript:

- Threshold procedure will be described step by step, with equations, the number of events per duration, and parameter uncertainty.
- We will state how many of the 104 events have a reported time of occurrence and how many were assigned a time using the peak-rainfall rule.
- The same timing rule will be applied to the non-triggering windows, so that both samples are treated symmetrically.
- A sensitivity analysis will re-derive the threshold under alternative timing assumptions (end of the reported day; hour of peak rainfall; timed events only) and report the effect on α, β and the validation scores.

Comment 1.5. *The runout and building-exposure modules are demonstrated for one event but not quantitatively validated. Building exposure alone should not be described as impact or risk.*

Response: We agree. Risk requires hazard probability, exposure and vulnerability, and ALERT currently estimates only rainfall-triggering hazard and the exposure of buildings. We will correct the terminology throughout and add a quantitative evaluation of the runout and exposure outputs for the Wayanad event.

Planned changes in the revised manuscript:

- Terminology: "impact-based warning", "risk", "buildings at risk", "assets under immediate threat" and "successful prediction of the runout zones" will be replaced by "exposure", "exposed buildings" and "modelled runout extent". We will state explicitly that, despite the acronym, ALERT estimates hazard and exposure, not risk in the strict sense.
- Runout evaluation for the 30 July 2024 Wayanad (Mundakkai–Chooralmala) debris flow against the mapped event footprint (Yunus et al., 2025): area-based scores (true-positive rate, false-positive area, overlap index) and modelled versus observed runout length.

Comment 1.6. *The manuscript and released code do not fully agree regarding rainfall durations, antecedent rainfall, warning logic, and susceptibility constraints. The exact model used to generate the results must be clearly documented and archived.*

Response: We thank the reviewer for highlighting these inconsistencies. We recognize that consistency among the manuscript, the computational implementation, and the published results is essential for reproducibility. We will carefully review and reconcile the rainfall durations, antecedent rainfall calculations, warning logic, and susceptibility constraints implemented in the code with those described in the manuscript. We will also prepare a user manual describing the model configuration and execution procedure and provide an archived version of the code through GitHub, with a clearly identified release corresponding to the revised manuscript.

Comment 1.7. *Global transferability is claimed but not tested outside the Western Ghats. Providing a list of regional thresholds does not demonstrate that the system works globally.*

Response:

We agree. A catalogue of thresholds shows that the software can be configured for other regions; it does not show that the resulting warnings are skilful there. Also most thresholds in Table 1 were derived from rain-gauge records, whereas ALERT applies them to IMERG and ECMWF grid-cell rainfall. Because satellite and model products smooth peak intensities, gauge-based thresholds are not directly transferable to these inputs and may lead to missed warnings. To strengthen this aspect, we plan to compile additional dated landslide inventories from different

geographical regions and evaluate the applicability of the available regional and global rainfall thresholds wherever suitable data are available. We will also revise the manuscript to distinguish the framework's potential geographical transferability from its demonstrated predictive performance and to avoid unsupported claims of global validation.

Planned changes in the revised manuscript:

- "Globally transferable" and similar claims will be made moderate as per the new analysis. We will distinguish the framework's technical transferability (configurable thresholds and susceptibility layers) from its demonstrated performance, which is limited to the Western Ghats.
- Table 1 will gain columns for rainfall data source (gauge, radar or satellite) and duration range, with a caution on applying gauge-based thresholds to satellite rainfall.
- Out-of-region test: we will evaluate ALERT in at least one more additional region with a publicly available dated inventory, for example Italy using the ITALICA catalogue (Peruccacci et al., 2023, ESSD 15, 2863–2877), applying both the published Italian thresholds and an IMERG-recalibrated threshold.
- Section 6.3 will be revised accordingly.

> **Comment 1.8.** *There are also several inconsistencies: accuracy is exactly 0.70, not exceeding 0.70; 5,474 + 7,707 equals 13,181, not 13,188; the abstract reports 24 thresholds whereas the manuscript reports 23; and Figure 9 contains only 65 landslide samples despite stating that 104 were available.*

Response: We sincerely thank the reviewer for carefully identifying these discrepancies. We will systematically verify and correct numerical inconsistencies throughout the manuscript, including reported accuracy, total sample count, and the number of rainfall thresholds. Regarding Figure 9, the apparent discrepancy between the reported and visually identified landslide samples is due to overlapping data points. We will verify that all intended observations are represented, improve the graphical presentation as necessary, and clearly indicate overlapping observations and the total sample size in the revised figure caption. Other errors found during our check will also be corrected, including a duplicated sentence in Section 5.2, the reference to "Table 1" for the colour scheme (should be Table 2), and typographical errors in the equations.

We believe these revisions address the referee's concerns and will result in a more transparent and scientifically defensible manuscript. We thank Referee #1 again for the constructive review.

On behalf of all co-authors,
Ali P. Yunus