This paper presents a systematic analysis of debris-flow surge sequences based on observational data from Jiangjia Gully (JJG) and 12 global watersheds, and proposes a three-layer stochastic model to simulate surge sequence generation. The study reveals that surge magnitude and inter-surge intervals follow a power-law distribution with exponential cutoff and an exponential distribution, respectively. The model successfully reproduces key statistical characteristics of observed sequences. Overall, this research addresses an important gap in debris-flow research at the "sequence scale," demonstrating significant theoretical innovation and practical application potential. The dataset is substantial and the modeling approach is well conceived.

Recommendation
Minor Revision

Major Comments
1. The core assumption that source-area failures follow a Poisson process would benefit from additional supporting evidence

The manuscript assumes that landslide/failure events in source areas follow a Poisson process based on the exponential distribution of inter-failure intervals. This assumption is fundamental to the model framework. While the authors cite Guo et al. (2023) as experimental support, the description of this evidence is relatively brief.

Suggestions:
- In Section 3.4 or the Supplementary Information, expand the description of the experimental results from Guo et al. (2023), clarifying the experimental design, observation duration, and the number of failure events that support the Poisson nature of source failures;
- If the authors have access to direct monitoring data from JJG source areas (e.g., displacement sensors, micro-seismic records), consider adding a time-series plot for a representative period to visually demonstrate the random independence of failure events;
- If direct monitoring data are indeed limited, a brief acknowledgment in Section 5.5 would be helpful, noting that future work with dense seismic arrays or InSAR observations could further validate this assumption.

Reply: We have added a brief statement in Section 5.5 (Lines 869–872) acknowledging that the Poisson assumption rests primarily on Guo et al. (2023), direct monitoring remains limited, and future validation could benefit from dense seismic arrays or InSAR.

2. The physical meaning of the Gaussian selection coefficient ζ could be clarified further

The selection coefficient $\zeta \sim N(0,1)$ is introduced to represent the combined effects of intermediate processes (channel erosion, sediment entrainment, etc.). While mathematically elegant, readers may wonder whether this normal distribution is universally applicable across all watersheds and events, and whether its mean and variance should be fixed or site-dependent.

Suggestions:

- Consider adding a brief sensitivity analysis (potentially as Supplementary Material) showing how variations in the variance of ζ affect simulation outcomes. If results are relatively insensitive to ζ parameters, this would strengthen confidence in the simplification;
- In the Discussion, the authors could note that, with more comprehensive data, the distribution parameters of ζ could potentially be linked to watershed characteristics such as drainage area or main-channel slope. This would acknowledge the current simplification while presenting it as a direction for future development.

Reply: We have added a brief discussion in Section 5.5 (Lines 872-875) acknowledging that $\zeta \sim N(0,1)$ is a simplification justified by the Central Limit Theorem, while noting that site-specific variations may exist. We also state that a systematic sensitivity analysis and site-dependent calibration of ζ will be pursued in future work as more observational data become available.

3. Model validation could be strengthened by examining temporal structure

The current validation relies primarily on the KS test (distributional similarity) and visual comparison of the moving average decay trend. The KS test does not assess temporal correlations (e.g., autocorrelation structure), yet the manuscript notes in Section 3.2 that surge sequences exhibit significant autocorrelation and high Hurst exponents. Examining whether simulated sequences also reproduce these temporal structures would strengthen the validation.

Suggestions:

- Consider adding a simple additional validation metric: compute and compare the first several autocorrelation coefficients (e.g., lag-1 to lag-10) or Hurst exponent estimates between observed and simulated sequences;
- This would not require any modification to the model itself, only an additional post-hoc evaluation. If the simulated sequences match the observations on these metrics, it would further demonstrate model effectiveness; if discrepancies exist, candidly discussing them would still provide valuable insight.

Reply: We agree that examining temporal correlations provides a more complete validation of the model. Following this suggestion, we have added a brief temporal structure validation in Section 4.1 (Lines 744-746). The Hurst exponents of the simulated sequences range from 0.65 to 0.89 (mean $\pm$ SD: 0.76 ± 0.09), nearly identical to the observed sequences (0.65–0.89). The autocorrelation functions for lags 1–10 also show close agreement between simulated and observed sequences, as shown in the newly added Figure S7 (Lines 316-320), with Hurst exponents annotated in the legend. This additional validation confirms that our stochastic model captures not only the magnitude distribution but also the temporal persistence of the surge sequences. We believe this concise addition addresses the reviewer’s suggestion without overburdening the main text.

4. The scope of model applicability could be more clearly defined

The cross-watershed validation using Chalk Cliffs and Gadria Creek is encouraging. However, these sites have relatively short surge sequences (<50 surges) and, like JJG, are primarily "source-failure-dominated" debris flows. Readers may naturally wonder whether the model applies to "runoff-dominated" debris flows (e.g., Cancia in Italy or post-fire sites in the southern US).

Suggestions:

The authors need not conduct extensive new validation for such sites within a minor revision; however, a brief addition to Section 5.5 (Limitations) would be valuable, clearly stating that the model is best suited for cases where source-area failures are the primary sediment source. For runoff-erosion-dominated debris flows, additional hydrological modules may be required. This would honestly delineate the model's applicability and point toward productive future work.

Reply: Following the reviewer’s suggestion, we have revised Section 5.5 (Lines 865-868) to clarify that our framework is best suited for source-failure-dominated catchments, while runoff-erosion-dominated settings would require additional hydrological modules.

Minor Comments

1. Terminology and symbols could be unified

- “Sediment delivery” and “sediment volume” are used interchangeably throughout the manuscript to refer to the same physical quantity *S*. It is recommended to adopt a single term consistently— “sediment volume” may be more appropriate since *S* represents a volume —or clearly state at first use that the two terms are used synonymously.
- The symbol ζ is used both as the Pareto shape parameter (near Eq. 8) and as the selection coefficient (Eqs. 10-11). While context generally disambiguates, it is advisable to change the Pareto exponent to ξ and keep ζ for the selection coefficient to avoid potential confusion.

Reply: We have replaced “sediment delivery” with “sediment volume” throughout the manuscript for consistency. Regarding the symbols, we confirm that ξ and ζ have been used consistently throughout (ξ for the Pareto exponent and ζ for the selection coefficient), so no revision was needed.

2. A few language expressions could be slightly refined

- A few long sentences in the Abstract and Introduction could be broken down for improved readability. For example, lines 11-14 (“While previous studies... remain poorly understood”) would benefit from being split into two sentences.
- The sentence “this approach prioritizes operational feasibility over detailed mechanistic understanding” (end of Section 3.4) is somewhat colloquial. Consider rephrasing to "this approach trades detailed mechanistic representation for operational feasibility" or a similar more formal expression.

Reply: Following the reviewer’s suggestion, we have split several long sentences in the Abstract (Lines 12–13) and Introduction (Lines 58–59, 62) to improve readability. We have also rephrased the colloquial expression in Section 3.4 (Lines 570 – 571) to “trades detailed mechanistic representation for operational feasibility.” We believe these revisions improve the clarity of the manuscript.

3. Brief explanatory notes for the tables in Supplementary Information would be helpful

Table S5 lists simulation parameters for global watersheds but does not include a note explaining the physical meaning of each parameter (particularly the α values) and their potential relationship with watershed characteristics. Adding one or two sentences in the table note — for instance, identifying possible controlling factors that differentiate sites with larger α (e.g., CC) from those with smaller α (e.g., GC) — would help readers better interpret the cross-watershed variation in parameters.

Reply: We have added a footnote to Table S5 (Lines 357-362) in the Supplementary Information explaining that α is the Pareto exponent describing the slope of the failure-mass distribution, with larger values indicating a steeper distribution (fewer large failures) and smaller values a heavier tail (more frequent large failures). We also note that cross-watershed variation in α primarily reflects differences in lithology, weathering, and source geometry. We believe this note helps readers better interpret the parameter variations across sites.

Additional Comments

1. Practical guidance for hazard warning could be slightly expanded

Section 5.4 discusses the model's application to hazard assessment, but the description remains rather general. It would be helpful for engineering practitioners if the authors could add one or two sentences specifying, in a real-time warning context, what input data the model would require (e.g., rainfall intensity, antecedent soil moisture) and what the approximate prediction time window would be.

Reply: We have added two sentences at the end of Section 5.4 (Lines 852–856) specifying that the model requires rainfall intensity and source-area mapping as inputs for real-time warning, with antecedent soil moisture further constraining λ. We also note that the model can provide forecasts minutes to hours ahead of arrival, depending on watershed scale. We believe this addition enhances the practical utility of our framework for hazard warning.

2. The theoretical significance of the power-law with exponential cutoff deserves stronger emphasis

The distribution given by Eq. 3 — a power law with an exponential cutoff — is an important theoretical contribution of this work. The authors may wish to highlight more explicitly in the Conclusions that this distribution contrasts with the pure power laws commonly observed for landslides and earthquakes, suggesting a "finite-size" characteristic of debris-flow systems (limited by the total available erodible material within the watershed). This may have broader implications for understanding the scale behavior of mountain hazard systems.

Reply: We have added a sentence (Lines 931-933) in the Conclusions emphasizing that the truncated distribution contrasts with pure power laws and reflects the finite sediment supply that constrains debris-flow magnitudes at the watershed scale.

Conclusion

This is a well-conceived and valuable study, supported by substantial observational data and a clear modeling framework. The authors address an important gap in the debris-flow literature by focusing on surge sequences as integrated systems rather than isolated events. The statistical findings are robust, the model is innovative, and the cross-watershed validation enhances the generalizability of the conclusions. The comments raised above are primarily aimed at strengthening the physical justification of key assumptions, expanding the validation metrics, and improving presentation clarity. They do not challenge the core methodology or main conclusions. I look forward to seeing the revised manuscript and believe that with these refinements it will be an excellent contribution to NHESS.

Reply: Thanks.