This manuscript presents a novel stochastic framework for modeling debris-flow surge sequences, addressing a significant gap in the literature. The authors compile an impressive global dataset, identify universal statistical characteristics across diverse watersheds, and develop a parsimonious model that successfully reproduces observed sequence statistics. The discovery of a unified magnitude–frequency distribution represents an important theoretical advance. The methodology is robust, the validation is thorough, and the potential for operational hazard forecasting is clearly demonstrated. Overall, this work makes a substantial contribution to probabilistic hazard assessment and is well-suited for publication in NHESS pending minor revisions.

Comments

1. The KS test results are inconsistent. The authors claim that $D < D_{0.01}$ for all events, but in the Supplementary Material, event 000704 has a *p*-value of 0.004, which would reject the null hypothesis at $\alpha = 0.01$.

Reply: Following your advice, We have re-checked Table S2 and confirm that for event 000704, $D = 0.300 > D_{0.01} = 0.195$ and $p = 0.004$, so the null hypothesis is rejected at $\alpha = 0.01$. The original statement "$D < D_{0.01}$ across all cases" was inaccurate.

We have made the following minimal revisions:

(1) Section 4.1: changed "across all cases" to "for six of the seven analyzed events", and added a brief clause noting that event 000704 is an exception ($D = 0.300$, $p = 0.004$) attributed to its long sequence and pronounced high-magnitude fluctuations not fully captured by the simplified inter-process representation. Also changed "*p*-values substantially exceed 0.05" to "for the other six events, the *p*-values exceed 0.01". (Lines 672-680).

(2) Section 4.2.1: limited "all events" to "all CC and GC events" to avoid conflict with the JJG exception (Lines 749-750).

(3) Abstract and Conclusions: added the qualifier "for the majority of analyzed events" to avoid overstatement (Line 26 and 935).

(4) Table S2: added a one-line note stating $D>D_{0.01}$ and $p = 0.004$ for event 000704 (Line 330).

We believe these minimal changes resolve the inconsistency while preserving the overall conclusions. We thank the reviewer again for this careful check.

2. The description of simulation results in the main text is overly absolute. The authors state that "simulated results show no significant difference from observed data," but in the Supplementary Material, event 000704 (p = 0.004) and event 990716 (p = 0.075) exhibit fit quality near or below the significance threshold, indicating inconsistency between the main text and the Supplementary Material.

Reply: We agree that the description of the simulation results in the original manuscript was overly absolute, and that events 000704 (p = 0.004) and 990716 (p = 0.075) exhibit fit quality near or below the significance threshold. We have revised both the main text and the Supplementary Information to present the validation results more accurately and transparently.

(1) Section 4.1 (main text): We have replaced absolute statements such as "no significant difference" and "strong agreement" with more measured wording ("generally agrees well", "reasonable agreement"). We now explicitly state that "$D < D_{0.01}$ for six of the seven analyzed events" and identify event 000704 ($D = 0.300 > D_{0.01} = 0.195$, $p = 0.004$) and event 990716 ($p = 0.075$) as cases with fit quality near or below the significance threshold (Lines 665-667).

(2) Section 4.2.1 (main text): We have clarified that "all events" refers to "all CC and GC events" (Table S3), avoiding conflict with the JJG edge cases (Line 341).

(3) Table S2 (Supporting Information): We have added a note explicitly listing D, $D_{0.01}$, and p for events 000704 and 990716, and stating that they are near or below the significance threshold (Lines 330-331).

We believe these revisions resolve the inconsistency between the main text and the Supplementary Information, and present the model's performance honestly without overstating its accuracy.

3. The parameterization relating source failure to rainfall intensity lacks quantitative validation. The goodness-of-fit (e.g., R^2, RMSE, confidence intervals) for $\lambda = \exp(k_{\mathrm{T}} \cdot I_{\mathrm{R}})$ is not provided in the manuscript.

Reply: We acknowledge that the original manuscript presented the parameterization $\lambda=\exp(k_T \cdot I_R)$ without explicitly documenting its goodness-of-fit, and we apologize for this omission.

The exponential relationship between Poisson intensity and rainfall intensity is not a new assumption introduced in this study, but is derived from our previously published field experiments in Jiangjia Gully (Guo et al., 2021, 2023). In those experiments, seven sets of artificial rainfall tests were conducted on JJG source slopes under rainfall intensities ranging from 12 to 60 mm/h, and the failure rate was found to increase systematically with rainfall intensity, consistent with the exponential form.

To address the reviewer's concern, we have made the following revisions:

(1) Section 3.5 (main text): We have added explicit references to the experimental basis of the λ-I_R relationship, stating that it is supported by the seven rainfall experiments in JJG and citing Guo et al. (2021, 2023) (Lines 596-598).

(2) Supplementary Information: We have added a table summarizing the experimental conditions (rainfall intensities, failure counts, mean intervals) from Guo et al. (2021), enabling readers to verify the exponential trend. The original goodness-of-fit statistics are reported in the cited publication (Line 590).

4. The evidence for self-similarity is insufficient. The authors base their demonstration on a single event (E910709) divided into four subsequences; evidence from one event is inadequate to support such a universal conclusion.

Reply: We agree that demonstrating self-similarity on the basis of a single event (E910709) is insufficient to support a universal conclusion, and we have revised the manuscript accordingly. We acknowledge that a full multi-event test is beyond the scope of the present study. We have therefore revised the manuscript to present the E910709 analysis as a representative example rather than a universal proof. Specifically:

(1) In Section 3.2, we now state that "we present this analysis as a representative example rather than a universal proof; a systematic multi-event test is identified as a direction for future work" (Lines 322-324).

(2) In the summary of Section 3.2, we have added that "while our demonstration is based on a single representative event, the same statistical features are recovered in the global datasets (Section 4.2), supporting the view that this similarity is a general characteristic of surge sequences" (Lines 369-371).

(3) In the Abstract, we have changed "self-similar structure in sequence organization" to "a self-similar structure in sequence organization identified in a representative long sequence" (Line

19-20).

(4) In Section 5.1 and the Conclusions, we have added the qualifier "observed in long sequences" to avoid overstating universality (Lines 929-930).

(5) In Section 5.5 (Limitations), we have added a sentence noting that "the self-similarity analysis is currently demonstrated on a single long sequence; a systematic test across all available multi-surge events would further consolidate this finding" (Lines 902-904).

(6) In the Supporting Information, we have added a one-line note clarifying that the analysis is based on E910709 as a representative long sequence (Lines 84-86).

5. The application of the model to unmonitored watersheds lacks guidance on parameterization. For watersheds without historical debris-flow records, the manuscript does not provide clear operational guidance on how to estimate key parameters such as λ, m_0, and ζ.

Reply: We agree that the original manuscript lacked operational guidance for unmonitored watersheds. In Section 3.5, we have added a concise guideline: λ is estimated from the source area (DEM/remote sensing) and local rainfall intensity (regional records), with k_T adopted from nearby instrumented catchments; m_0 and ξ are obtained from regional landslide inventories or published ranges; and $\zeta \sim N(0,1)$ requires no site-specific calibration. We also note in Section 5.4 that a first-order simulation for an ungauged catchment requires only the source-area extent, local rainfall intensity, and an order-of-magnitude estimate of m_0 and ξ. We believe this provides the clear operational guidance requested by the reviewer while keeping the manuscript concise (Lines 618-625).

We further note in Section 5.4 that a first-order simulation for an ungauged catchment therefore requires only (i) the source-area extent from DEM/remote sensing, (ii) local rainfall intensity, and (iii) an order-of-magnitude estimate of m_0 and ξ from regional inventories. The model is thus transferable to ungauged catchments with minimal data requirements (Lines 878-882).

6. Notation is inconsistent. ζ is used for both the Pareto exponent and the Gaussian selection coefficient; the bold symbols $\boldsymbol{S_F(j)}$ and $\boldsymbol{M}$ in Eq. (10) are not explicitly defined in the text.

Reply: We have carefully checked the manuscript and made the following corrections.

(1) ζ used for both the Pareto exponent and the selection coefficient. We confirm that these are two distinct symbols in the manuscript: the Pareto exponent of the failure-mass distribution is denoted by ξ, whereas the Gaussian selection coefficient is denoted by ζ. The apparent inconsistency likely arose from the visual similarity between ξ and ζ in the PDF, and from the fact that the two symbols were not explicitly contrasted at their first occurrence. To eliminate any ambiguity, we have:

- Checked every occurrence of ξ and ζ throughout the main text and the Supporting Information, and corrected any misused instances;
- Added an explicit clarification at the first occurrence of ξ in Section 3.4: "where ξ is the Pareto exponent of the failure-mass distribution" (Line 442);
- Added a parenthetical clarification in Section 3.5: "the selection coefficient ζ (distinct from the Pareto exponent ξ)" (Line 608);
- Added both symbols to the Appendix A Notation table: ξ — exponent of the failure power-law (Pareto) distribution; ζ — selection coefficient, modeled as a standard normal variate $N(0,1)$ (Line 977).

(2) Undefined symbols in Eq. (10). We have added explicit definitions immediately following Eq. (10):

- $S_F(j)$ denotes the mass of the j-th surge in the final sequence in the main channel (Line 553);
- M is the mass vector from the source-failure sequence $\{S_0\}$ (Line 554).

We have also replaced the generic statement "where the symbols are the same as in Eq. 10" after Eq. (11) with explicit definitions of $S_F(j)$, $S_0(\tau(j))$, and $\zeta(j)$.

We believe these corrections fully resolve the notation inconsistencies and improve the clarity of the model formulation. We thank the reviewer for this careful reading.

7. Figures 6 and 8 contain too many subpanels, with small fonts and crowded legends, reducing the readability of key results. The statistical parameter comparison subpanels occupy substantial space while conveying limited information.

Reply: We have revised both figures accordingly. The font sizes of the axis labels, tick labels, and legends in Figures 6 and 8 have been enlarged (Line 706 and 760), and the legends have been simplified to reduce visual clutter. These changes make the core results—the sediment fluctuations, the decay trend of $\langle S \rangle_n$, the scaling distributions, and the sediment-volume comparison—substantially easier to read.

8. Line 97: "systematical" should be "systematic."

Reply: revised.

9. Line 129: "lead fish sampler" is unclear; a more standard term or an explanation is recommended.

Reply: we have revised it into "sampling pole" (Line 163).

10. Line 337: "the same" should be "similar."

Reply: revised.

11. Reference formatting is inconsistent: some references include DOIs while others do not; journal names are a mix of full names and abbreviations.

Reply:

12. Line 585–586: The critical value c(0.01) = 1.63 is given without specifying the sample size conditions to which it applies.

Reply: We have carefully checked and revised the reference lists in both the main text (Lines 1003-1249) and the Supporting Information (Lines 377-470), and have standardized them fully in accordance with the NHESS (Copernicus) reference style.

13. The term "thinning" is not clearly defined when first introduced.

Reply: We have added an explicit definition at its first occurrence in Section 3.4:

"··· this can be reduced to a thinning—a standard operation on a Poisson process in which each event is independently retained with a certain probability, so that the retained events themselves form a new Poisson process with reduced intensity (Kingman, 1993)—from $\{S_0\}$ to $\{S_F\}$, i.e., ···" (Lines 564-566)

We have also added the brief parenthetical clarification “thinning (probabilistic selection)” at the first mention in the Abstract (Line 22), to ensure that the meaning of the term is clear to readers throughout the manuscript.