Recommendation
Minor Revision

1. Introduction (last paragraph on page 2): The assertion that "slope failures constitute the dominant initiation mechanism" requires a more nuanced discussion of its relationship with other mechanisms

The manuscript states that "Field evidence from multiple watersheds definitively demonstrates that slope failures constitute the dominant initiation mechanism, rather than flow instabilities" and provides three lines of supporting evidence. This assertion directly underpins the model's physical basis — namely, that the Poisson process in source areas is the primary driver of surge generation.

However, this paragraph dismisses "flow instability" in a single sentence before reaching its conclusion. The manuscript also acknowledges that "flow instabilities may modulate subsequent surge propagation dynamics," but does not clarify the extent or manner of this "modulation." Readers may wonder: if slope failures are "dominant," is the "modulation" merely a minor correction, or could it become more significant under certain conditions (e.g., steeper main-channel gradients, higher sediment concentrations)?

Please add one or two sentences at the end of this paragraph summarizing the relative importance of slope-failure dominance versus flow-instability modulation, and how their relative contributions might vary across different watershed conditions (such as slope, concentration, and discharge). This does not require quantitative analysis but would help readers better understand the model's scope of applicability.

Reply: We agree that our original wording was overly absolute and did not adequately clarify the relationship between slope failures and flow instabilities. In the revised manuscript, we have added a brief clarification at the end of the paragraph (Lines 388-392) to emphasize that slope failures serve as the primary initiation mechanism governing surge timing and intermittency at the watershed scale—thus justifying our Poisson-process representation—while flow instabilities (e.g., roll waves) play a secondary modulating role that becomes more pronounced in steep, long channels with high sediment concentrations. Conversely, in catchments with mild slopes or short transport paths, the source-controlled Poisson process dominates the observed surge statistics. This revision, we believe, better delineates the scope and applicability of our stochastic framework without introducing quantitative analysis, as the reviewer suggested.

2. Section 2.1 (Data Source from Global Watersheds) and SI Text S2: The number of watersheds is inconsistent — the main text says 12, but SI Text S2 lists 15

The manuscript states in Section 2.1 that "additional datasets from 12 representative global watersheds were analyzed" and shows 12 sites in Figure 1b. However, Supplementary Information Text S2 lists 15 watersheds (from Illgraben to Rebaixader). Please verify and reconcile the numbers between the main text and the SI, clarifying which 12 were included in the quantitative analysis and whether the remaining ones are only described as background or are used in the event sequences shown in Figure 7b.

Reply: Following your advice, we have reconciled the numbers between the main text and the SI as follows. In the revised main text (Section 2.1, Lines 122-129), we now state that a suite of global watersheds was compiled (Figure 1b; details in Text S2), among which 12 catchments with sufficient surge sequences were used for quantitative event-sequence analyses; the additional catchments described in SI Text S2 provide complementary background information but were not included in the sequence-level statistical analyses due to limited event records. We have also revised the caption of Figure 1b to clarify that the 12 labeled catchments were those used in the quantitative analyses (Lines 153-155), and we have added a clarifying sentence at the beginning of SI Text S2 to explain the distinction (Lines 85-89). We believe these revisions resolve the inconsistency and improve the transparency of our data selection.

3. Section 2.3: The phrase "the progressing n can be substituted by the corresponding time of the surge" is somewhat ambiguous

This sentence appears in the text below Eq. (1). Consider rephrasing it for greater precision, for example: "the index n can be mapped to the chronological time of each surge using the cumulative time from the event onset," so that readers more easily understand the relationship between n and time.

Reply: Thanks for your suggestion, we have replaced "the progressing n can be substituted by the corresponding time of the surge" with "the index *n* can be mapped to the chronological time of each surge using the cumulative time from the event onset" in the revised manuscript (Lines 192-193). This revision improves the precision of the description.

4. Section 3.4 (Mathematical Formulation), Eq. (9): There is a tension between the mathematical expression of water-sediment coupling and the observational evidence

Eq. (9) expresses total surge discharge as the superposition of water discharge $Q_w(t)$ and discrete sediment supply terms M(t): $Q_\mathrm{f}(t) = Q_\mathrm{w}(t) + \sum_{t_i<t} M(t)\delta(t-t_i)$. This is mathematically clear. However, Figures 5a and 5b show that the observed surge sequences and the simulated SCS hydrographs are not temporally synchronized — the hydrograph peaks often occur after the surge sequences have ended.

If $Q_w(t)$ and $M(t)$ are not synchronized in time, how are their timescales coordinated in Eq. (9)? When $Q_w(t)$ reaches its peak, $M(t)$ may have already decayed to zero, and vice versa. In such cases, does simple superposition produce a total discharge $Q_f(t)$ that lacks a clear physical temporal correspondence?

Please provide some clarification on this issue: how is the temporal alignment between $Q_w(t)$ and M(t) handled in model implementation, given that Figure 5 demonstrates their asynchrony? Is Eq.

(9) merely a "conceptual" expression, with the role of $Q_w(t)$ in practice integrated into the Gaussian selection coefficient (as shown in Eqs. 10-11)? If the latter is the case, please state this explicitly to avoid confusion regarding the relationship between Eq. (9) and the subsequent Eqs. (10-11).

Reply: We agree with the reviewer that the relationship between Eq. (9) and the subsequent model implementation (Eqs. 10–11) was unclear, and that the asynchrony shown in Figure 5 precludes a direct deterministic superposition. To address this, we have revised Section 3.4 (493–521) to distinguish the conceptual formulation from the operational model. Equation (9) is now presented as a conceptual reference for the physical components of water–sediment coupling, rather than as a time-stepping equation. Given the observed hydrograph–surge asynchrony, we do not prescribe $Q_w(t)$ explicitly. Instead, the net effects of hydrological processes are integrated into a Gaussian selection coefficient $\zeta \sim N(0,1)$, which probabilistically governs mass contributions to surges — this is the mechanism implemented in Eqs. (10)–(11) and the thinning procedure. A concise summary has also been added after Eq. (11) (557–559) to reinforce this distinction. We believe these revisions eliminate the ambiguity and clarify the logic from conceptual physics to model implementation.

5. Section 3.2 (Universal Statistical Characteristics): The reported range of the power-law exponent q could be more precise

The text states "with q ranging from 0.24 to 0.98" but does not specify how many events this range is based on, nor does it provide the mean or standard deviation of q. Please add the sample size and measures of central tendency to help readers assess the representativeness of the q values.

Reply: We agree that providing only the range of q was insufficient. In the revised manuscript, we have supplemented the description with the sample size, mean, and standard deviation of q: "with q ranging from 0.24 to 0.98 across 42 events (mean = 0.56, SD = 0.19)." (Line 265) We believe this additional information allows readers to better assess the representativeness and variability of the observed decay exponents.

6. Sections 4.1 and 4.2: The statistical moment comparisons for JJG and global watersheds overlap in content

In Section 4.1, Figure 6c-2 compares statistical moments (mean, median, standard deviation, range, skewness, kurtosis) between observed and simulated sequences for JJG events, concluding that "All metrics are comparable." In Section 4.2.2, Figure 8c-2 applies the identical statistical moment comparison to CC and GC events, arriving at the same conclusion. These two sections are highly repetitive in methodology and conclusions, with insufficient differentiation.

Consider moving the JJG statistical moment comparison (Figure 6c-2) to the Supplementary Information, retaining only the KS test and decay trend results in the main text, and using the additional space to highlight the differences between JJG and CC/GC (such as how sequence length affects validation strategy). Alternatively, if both sets of moment comparisons remain in the main text, please at least clarify in the text that they serve distinct purposes — for example, the former demonstrates model stability for long sequences, while the latter demonstrates model effectiveness for short sequences.

Reply: Thanks for the reviewer's suggestion. To differentiate the two moment comparisons, we have revised the text to clarify that the JJG analysis (Fig. 6c-2) validates the model for long sequences (hundreds of surges), whereas the CC/GC analysis (Fig. 8c-2) provides complementary evidence for short sequences (10–50 surges), where distribution tests alone are less conclusive. This distinction is now explicitly stated in Sections 4.1 (Lines 667-670) and 4.2.2 (Lines 724-726), respectively, without relocating any figures.

7. Section 5.1 (Generic Implications of Surge Sequences): The discussion of the mechanisms generating self-similar surge structures could be deepened

The manuscript notes that long-sequence events (such as E910709) can be divided into multiple sub-events (Groups A, B, C, D), each exhibiting the same statistical characteristics as the overall sequence (Figure 3b). This is important evidence supporting the "cascading thinning" model. However, the explanation for how self-similar structures arise is relatively brief, stating only that "mass supplies are randomly distributed in space and time" and linking this to "superposition of materials derived from multiple sources."

Please add one or two sentences at the end of this section discussing whether self-similarity implies that sub-watersheds at different spatial scales share similar statistical properties of source-area failures (i.e., scale invariance), or whether it reflects a "filtering" or "renormalization" effect of the main channel on sediment supplies from different sources. This would more closely link the observational phenomenon to the core mechanism of the model (cascading thinning).

Reply: We agree that the explanation of self-similarity in the original manuscript was overly brief. In the revised manuscript, we have added two sentences at the end of Section 5.1 to discuss the origin of self-similarity (Lines 776-779). We now explicitly state that self-similarity reflects both (i) the scale-invariant statistical properties of source-area failures across sub-watersheds of different sizes, consistent with the power-law distribution of failure masses, and (ii) the hierarchical filtering effect of the cascading thinning process (Eqs. 10–11), which renormalizes sediment inputs from multiple sources into surges that preserve the same statistical structure at each level of the watershed network. This addition directly links the observational phenomenon to the core mechanism of our model.

8. Section 5.4 (Application in Hazard Prediction and Assessment): The statement about clustering of distribution parameters could be more precise

The text states "(μ, M_c) clustering near μ = 0 indicates that large-magnitude surges follow exponential distribution." However, at what spatial scale was this clustering observed? Is it a global clustering across all studied watersheds, or clustering within individual watersheds? Please add qualifiers such as "across all studied watersheds" or "within individual watersheds" to clarify.

Reply: In the revised manuscript, we have added the qualifier "across all studied watersheds" (Line

839) to specify that the clustering of (μ, M_c) near $\mu = 0$ is a global pattern observed in the combined dataset from all catchments analyzed.

9. Section 5.5 (Limitations and Perspectives): Does the model's neglect of GSD evolution affect its predictive accuracy for sediment volume *S*?

The manuscript acknowledges that the model neglects "grain-size evolution during transport" and notes that the "ρ-GSD relationship" and "rheology variations dependent on GSD" are directions for future incorporation. However, sediment volume *S* is defined as the "volume of solid material," and solid particles of different grain sizes exhibit different settling velocities and erosion capacities during transport. Does the model's prediction of *S* without distinguishing grain sizes affect its applicability across different watersheds or events?

Please add a paragraph in this section assessing the potential error range or applicability constraints introduced by neglecting GSD evolution within the current model framework. For example, is this simplification more reasonable for watersheds with relatively uniform source grain-size distributions, whereas more detailed characterization is needed for watersheds with highly heterogeneous distributions? This would help readers judge the model's credibility under different geological conditions.

Reply: We have added the following sentence to Section 5.5 (Lines 864–867): "For catchments with relatively uniform source grain-size distributions, treating S as a bulk quantity is reasonable; however, for highly heterogeneous settings, the predicted *S* may deviate from actual transported volumes, particularly where size segregation during long-distance transport is pronounced (Wang et al., 2018; Yang et al., 2019)." This clarifies that our treatment of *S* as a bulk quantity is justified for uniform source materials, while acknowledging its limitations in heterogeneous settings where size segregation affects transport.

10. Figure 7b: The caption is potentially ambiguous

Figure 7b shows "Debris flow events from various locations," with the caption listing 12 locations including Illgraben, Gila National Forest, Nevada, etc. The vertical axis is labeled "Q (m^3/s)" and the horizontal axis "Event index," with multiple "event sequences" shown simultaneously. Please clarify whether this "event sequence" refers to historical events ordered chronologically within each watershed, or whether events from all 12 watersheds are merged and sorted in some manner. The current caption could be interpreted in either way.

Reply: We have revised the caption of Figure 7b to clarify that each color-coded sequence represents historical events ordered chronologically within the respective watershed; the event index thus denotes the temporal rank of each event in that catchment (Lines 716–718).

11. Supplementary Information Table S5: Dimensional consistency of parameter m_0

Table S5 lists m_0 with units of "10^{-3} m^3," with values for CC and GC ranging from 10.2 to 184.8, whereas Table S1 shows JJG's m_0 ranging from 5.02 to 5.25. The magnitude of m_0 differs by several to dozens of times between the two tables. Please add a note to Table S5 explaining the physical meaning of m_0 uniformly and briefly explaining the possible reasons for cross-watershed parameter differences (such as differences in source area, lithology, and degree of weathering). This would prevent reader confusion about the comparability of the parameters.

Reply: We have added a footnote to Table S5 (Lines 337-340) explaining that m_0 is the Pareto lower-cutoff scale parameter, and its variation primarily reflects differences in lithology, weathering, and source geometry, with larger values for CC and GC consistent with coarser source materials.

12. Supplementary Information Figures S4 and S6: Titles need to be clearly distinguished

Both figures currently have the identical title "Illustrating comparison of moving average decay between simulated and observed surge sequences for six debris-flow events." Please specify the event sets for each figure (e.g., S4 for JJG events, S6 for CC and GC events) and ensure that the main text citations point to the correct event sets to avoid reader confusion.

Reply: We have revised the titles of Figures S4 (Lines 291-293) and S6 (Lines 308-311) to clearly distinguish the event sets: Figure S4 now refers to JJG events and Figure S6 to CC and GC events. We have also verified that the main text citations point to the correct figures.