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 $\alpha = 0.01$.

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.

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

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.

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.

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

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.

8. Line 97: “systematical” should be “systematic.”

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

10. Line 337: “the same” should be “similar.”

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

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

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