Response to Referee #1

In this response, the review comments are stated first (in black), followed by our "Replies" (in blue).

Referee #1:

I find the manuscript interesting and potentially important. The dataset is broad, and the LPPLS-based procedure appears internally coherent for systems that exhibit accelerating, intermittent precursory dynamics. However, in my view, the current evidence supports a suggestive empirical pattern more strongly than a universal physical law. My main concern is that the central scaling may depend on model-derived onset times extracted from a retrospective compilation that may be conditionally selected toward events with analysable precursory acceleration. The manuscript would be substantially strengthened by clearer inclusion/exclusion criteria, sensitivity tests to onset definition and model assumptions, comparisons with simpler null models, and a more cautious interpretation of the finite-size criticality claim.

Reply: We thank the referee for this balanced and constructive assessment.

Our dataset was assembled by compiling, as extensively as possible, cases of landslides, rockbursts, glacier breakoffs, and volcanic eruptions reported in the literature and public repositories for which quantitative precursory time series could be obtained. Within the compiled candidate dataset, exclusions were made primarily on the basis of data adequacy, most notably when fewer than 15 data points were available during the pre-failure acceleration phase, which is insufficient for a reliable LPPLS analysis. We will revise the manuscript to state these inclusion/exclusion criteria more explicitly and to clarify that the compilation is extensive but not exhaustive.

At the same time, we acknowledge that any retrospective literature-based compilation is inevitably conditioned by data availability and monitoring quality. In particular, events without sufficiently resolved precursory observations, or for which the relevant time series have not been published, cannot be evaluated using the present procedure. Our conclusions therefore apply to the analysed population of monitored events with sufficiently resolved precursory dynamics, rather than to all events without qualification.

Concerning the determination of t_1, we respectfully disagree that a model-derived onset is necessarily a weakness. A central clarification is that t_1 is not intended to represent an externally observed timestamp in the same sense as the final failure time t_c. Rather, it represents the inferred beginning of the dynamical regime over which the LPPLS description becomes applicable. The central purpose of the Lagrange regularisation approach is precisely to avoid prescribing this onset through visual inspection, heuristic thresholds, or another criterion disconnected from the model used to describe the subsequent acceleration. If the precursory phase is defined as the regime governed by LPPLS dynamics, the most internally consistent procedure is to let the model determine the time interval over which those dynamics are supported by the data. Alternative procedures would introduce separate criteria that may not be directly aligned with the LPPLS dynamics.

We agree that the resulting onset is model-conditioned and have revised the manuscript to avoid suggesting that t_1 is objective in an absolute, model-independent sense. However, such model conditioning is inherent in any attempt to identify the beginning of a precursory regime. Visual, threshold-based, acceleration-based, and change-point definitions also embody assumptions about what constitutes a meaningful onset, even when those assumptions remain implicit. The advantage of the Lagrange regularisation approach is that its assumptions are explicit, reproducible, and directly connected to the dynamical description being evaluated. We will clarify this conceptual basis and the associated limitation in the revised manuscript.

Finally, we will revise the abstract, discussion, and title to avoid presenting the observed scaling as a uniquely established universal physical law. The revised title will be "General scaling of precursory duration with event size across mechanically driven geohazards". We will state more cautiously that the extensive dataset reveals a broad empirical scaling among the analysed events and that this pattern is consistent with, but does not by itself uniquely establish, a finite-size criticality interpretation.

More specific comments:

The main concern is that the applicability of LPPLS procedure is conditional. The method can be expected to work primarily for events that already display sufficiently clear accelerating and intermittent precursory dynamics. The manuscript should therefore clarify whether the proposed scaling applies to mechanically driven geohazards in general, or only to monitored events showing well-developed LPPLS-like behaviour. This requires explicit inclusion/exclusion criteria and a transparent accounting of candidate events or time series that were considered but rejected.

Reply: We thank the referee for raising this important point, but we would like to clarify that the applicability of the LPPLS procedure is somewhat broader than suggested in the comment. The LPPLS framework is not restricted a priori to time series that already exhibit accelerating precursory dynamics. Rather, the nature of the dynamics is inferred from the fitted parameters. In particular, the exponent m characterises the degree of acceleration: values $m < 1$ correspond to accelerating behaviour, whereas $m = 1$ corresponds to a linear trend and $m > 1$ to decelerating behaviour. Thus, the procedure itself provides a way to assess whether a candidate time series is compatible with an accelerating precursory regime, rather than requiring such a regime to be assumed beforehand.

Our intended claim is therefore not conditioned on a prior visual or qualitative selection of events displaying well-developed LPPLS-like behaviour. We consider monitored mechanically driven geohazards for which sufficiently resolved quantitative precursory time series and event-size estimates are available, and use the LPPLS analysis to characterise the temporal dynamics of those precursors. The scaling reported in the manuscript concerns the duration of the inferred precursory regime for the events retained in the analysis.

Our dataset was assembled by compiling, as extensively as possible, cases of landslides, rockbursts, glacier breakoffs, and volcanic eruptions reported in the literature and public repositories for which quantitative precursory time series and event-size estimates could be obtained. Exclusions were made primarily on the basis of data adequacy rather than on whether the time series displayed an accelerating pattern. In particular, cases with fewer than 15 data points during the relevant pre-event interval were excluded because they do not provide sufficient information for a reliable LPPLS calibration. We will make these inclusion and exclusion criteria more explicit in the revised manuscript and provide a transparent accounting of the candidate cases considered.

We will also revise the manuscript to clarify the scope of the proposed scaling. The LPPLS framework does not impose acceleration by construction; instead, it allows the data to determine whether the inferred dynamics are accelerating, approximately linear, or decelerating through the fitted value of m. This is an important feature of the approach, as it avoids defining the onset of a precursor through an externally imposed heuristic criterion and instead provides a consistent framework for identifying and characterising the temporal evolution preceding failure.

The key variable, precursory duration, is not independently observed. It is inferred by optimizing the LPPLS fit over variable calibration windows. Therefore, the reported scaling depends directly on a model-derived quantity.

The claim that the onset is "objective" should be softened: it is objective only conditional on the LPPLS model class, parameter bounds, regularization scheme, and preprocessing choices. The authors should test whether the scaling persists under onset definitions independent of LPPLS.

Reply: We thank the referee for raising this point, which allows us to clarify an important methodological aspect of our approach. We agree that the precursory duration is not a directly measured observable in the same sense as displacement or seismic amplitude. Within our framework, the onset time t_1, and hence the precursory duration, is inferred through calibration of the LPPLS model. We do not, however, regard this model dependence as a limitation specific to our approach, nor do we think that an onset definition constructed independently of LPPLS would, for that reason alone, be more objective.

Scientific inference is generally model-mediated. Many quantities routinely used in the physical sciences, such as characteristic times, transition points, correlation lengths, relaxation times, source parameters, critical exponents, are not simply read directly from raw observations, but are defined and inferred within theoretical or statistical frameworks. This is a longstanding point in the philosophy of science, from Duhem's emphasis on the role of theoretical assumptions in empirical testing, through Hanson and Kuhn's discussions of the theory-dependence of observation, to Einstein's insistence that scientific concepts are not simply given by experience but provide the conceptual structure through which observations acquire physical meaning. The relevant question is therefore not whether an inferred quantity is model-dependent, but whether the assumptions defining it are explicit, physically motivated, reproducible, and empirically testable.

Any operational definition of the beginning of a precursor necessarily involves such assumptions. A displacement threshold presupposes a relevant amplitude scale; a velocity or acceleration criterion depends on differentiation and smoothing choices; a change-point method assumes a statistical structure for the regimes before and after the transition; and visual identification relies on an implicit phenomenological model. These procedures are not model-independent simply because their assumptions are external to the dynamical model used subsequently.

The central value of the LPPLS procedure is precisely that it does not treat t_1 as an externally prescribed breakpoint. The physical hypothesis underlying the method is that the system may enter a dynamical regime in which increasingly strong positive feedbacks organise the evolution towards terminal failure. If the onset of this positive feedback-dominated regime corresponds to a genuine physical change in the system, then its location should not be determined independently of the dynamics that follow. On the contrary, the beginning of the regime and its subsequent evolution should constrain one another.

This is the reason for determining t_1 endogenously within the LPPLS calibration. The onset is inferred jointly with the parameters describing the subsequent dynamics, so that the entire precursory trajectory contributes to identifying when the positive feedback-controlled regime begins. In our view, this is not a drawback of the procedure but one of its main strengths. It establishes a direct connection between the inferred beginning of the precursor and the dynamical structure of the approach to failure, rather than separating the two through an externally imposed threshold or independently selected breakpoint.

The resulting precursory duration therefore has a precise operational and physical meaning: it is the characteristic time interval over which the observed dynamics are most consistently described by the LPPLS representation of the approach towards failure. The endogenisation of t_1 is especially important in this respect. If the transition into the positive feedback-controlled regime is physically meaningful, then one would expect its identification to be constrained by the full temporal organisation of the precursor. LPPLS provides a systematic way to make exactly this connection.

We will nevertheless revise the manuscript to avoid any wording that could be interpreted as implying a model-independent onset. In particular, expressions such as "objectively identified onset" will be replaced by formulations such as "endogenously and reproducibly inferred onset within the LPPLS framework." Our use of the term "objectivity" was intended to refer to the explicit and reproducible nature of the inference procedure, rather than to imply independence from the underlying model class. We will therefore revise the terminology throughout the manuscript to make this distinction explicit.

We agree that the robustness of the inferred onset within the LPPLS framework should be assessed. This issue is, however, distinct from requiring an onset definition independent of LPPLS. The relevant robustness question is whether the LPPLS–Lagrange inference yields a well-constrained and reproducible onset rather than behaving as a flexible fitting procedure capable of generating arbitrary onset times. The Lagrange regularisation procedure already provides a means of assessing this. In particular, the analysis illustrated in Supplementary Fig. S2 shows that, as progressively more of the precursory record becomes available, the inferred converges towards a stable plateau rather than wandering arbitrarily with the calibration window. Once sufficient information about the positive feedback regime is present, the model therefore recovers a persistent estimate of its beginning. In response to the referee's comment, we will strengthen the presentation of this analysis in the revised manuscript and Supplement and make its implications for the robustness of the inferred precursory durations more explicit.

More generally, our claim is not that LPPLS uncovers a uniquely defined onset existing independently of any representation of the data. Rather, LPPLS is designed to identify the emergence of a particular dynamical regime associated with positive feedbacks that can drive the system towards terminal failure. The inferred t_1 is therefore part of the physical content of the model: it marks the beginning of the regime whose subsequent evolution the model seeks to describe.

Our central result is that the characteristic duration of this dynamically inferred regime, determined consistently using the same explicit framework across a broad range of mechanically driven systems, exhibits a systematic relation with event size. The fact that this timescale is model-derived does not weaken the result. On the contrary, it gives the timescale a precise dynamical interpretation, connects the inferred beginning of the precursor to the subsequent approach to failure, and makes the assumptions underlying the cross-system comparison explicit, reproducible, and falsifiable.

A related concern is possible model imprinting. The same model family is used to define the precursory duration and to support the physical interpretation of accelerating, oscillatory critical behaviour. This could make heterogeneous systems appear more comparable than they actually are. The authors should compare LPPLS-derived onset times with other onset criteria.

Reply: We understand the concern, but we do not think that using the same model family to identify the onset of the precursory regime and to interpret its subsequent dynamics constitutes model imprinting in the problematic sense implied here. On the contrary, this coupling is central to the physical logic of the LPPLS approach.

The LPPLS framework embodies a specific dynamical hypothesis: that, prior to failure, the system may enter a regime increasingly governed by positive feedbacks, producing an accelerating evolution that can be accompanied by intermittent, approximately log-periodic oscillation. Within this framework, the onset time t_1 is not an externally imposed breakpoint. It is inferred endogenously as the beginning of the interval over which the subsequent dynamics are most consistently described by that hypothesis. If the transition into this positive feedback-dominated regime is a genuine physical process, then its beginning should be constrained by the dynamics that follow. Estimating t_1 together with the parameters describing the approach to failure is therefore not circular; it is precisely what allows the inferred onset to be dynamically connected to the full precursory evolution.

This point is important because any alternative onset criterion is itself model-dependent. A fixed displacement threshold assumes a relevant amplitude scale; a velocity or acceleration threshold depends on smoothing and differentiation choices; a change-point method assumes a statistical model for the regimes on either side of the transition; and visual identification introduces an implicit phenomenological model. An onset estimate obtained "independently" of LPPLS is therefore not model-independent. It simply relies on a different set of assumptions, which are not automatically more objective or more physically meaningful. Moreover, these alternative assumptions are likely to be not fully compatible or in contradiction with the physics embodied in the LPPLS model.

The role of the Lagrange regularisation method is precisely to make the determination of the calibration window systematic and to avoid the trivial preference for increasingly short windows. Rather than selecting t_1 by eye or prescribing it through an external threshold, the procedure asks when the LPPLS description becomes sufficiently informative over the evolving record to justify identifying the beginning of the corresponding dynamical regime. In this sense, the endogenisation of t_1 is a key value of the method: it connects the inferred beginning of the precursor directly to the dynamical structure of the approach to failure.

We also do not view the use of a common model across heterogeneous systems as artificially making them comparable. On the contrary, the justification for employing a common dynamical framework is that LPPLS is intended to capture generic ingredients that recur across a wide class of failure processes: nonlinear positive feedbacks that amplify departures from stability, inertial or memory effects that sustain the evolution once initiated, and nonlinear negative feedbacks that interact and compete with the destabilizing processes. These ingredients are not specific to any one geohazard type; they represent broad dynamical mechanisms by which complex systems can organize towards instability and terminal failure. LPPLS can be shown to be a minimum embodiment of these three mechanisms.

The purpose of the common model is therefore not to erase heterogeneity, but to test whether different systems share a common dynamical skeleton beneath their system-specific details. Cross-system comparison in physics is meaningful precisely when one can identify such generic mechanisms and express them through a common set of dynamical variables or parameters. In the present case, the substantive question is whether mechanically driven systems with very different materials, geometries, forcing conditions, and observational characteristics nevertheless display a common relationship between the duration of the inferred positive feedback-dominated regime and event size.

The use of the same model does not impose the existence of such a scaling. It merely provides a consistent operational definition of the quantity being compared across systems. The scaling could still be absent, strongly scattered, or differ systematically between geohazard types. Its emergence is therefore a nontrivial empirical result, and, if robust, supports the interpretation that the observed relationship reflects generic features of the underlying failure dynamics rather than idiosyncrasies of any particular system or dataset.

The universality claim appears stronger than warranted. Volcanic eruptions are included in the broad framing of the manuscript, but they do not show a comparable scaling and are excluded from the main mechanically driven relationship. The rockburst subset also appears statistically weak or inconclusive. These points should temper the cross-hazard generalization.

Reply: We agree that the universality claim should be expressed in a more balanced way. The manuscript already distinguishes between (i) the broad occurrence of LPPLS-like accelerating and intermittent dynamics across several geohazard classes and (ii) the specific volume–duration scaling observed for mechanically driven instabilities. It also explicitly reports that volcanic eruptions do not exhibit a statistically significant scaling with failure volume.

Regarding rockbursts, we agree that the sample size is limited and that the rockburst-only regression is statistically inconclusive. We will state this more explicitly in the revised manuscript. We present the rockburst data as compatible with the mechanically driven trend but not, by themselves, sufficient to establish a separate scaling. This distinction is important and is clarified.

We will also replace "universal scaling" with "general scaling" throughout the manuscript, including in the title.

It is surprising that a single-parameter scaling with volume works as well as reported, given that precursory duration should also depend on material properties, geometry, boundary conditions, monitoring resolution and others. The authors should probably examine residuals by hazard type, material class, monitoring method, temporal resolution, and data source, and test whether volume remains significant in a multivariate or hierarchical model where possible.

Reply: We agree that event volume cannot be the sole control on precursory duration. Material properties, geometry, loading and boundary conditions, as well as observational factors such as monitoring resolution and data quality, are all expected to influence the duration inferred for a given event. Our interpretation is therefore not that event volume V provides a complete predictive model for duration T, but that volume captures a leading-order finite-size dependence that becomes visible when events spanning many orders of magnitude are considered together.

A more general representation of the dependence would be of the form

$$T = V^a F(p_{mat}, p_{geom}, p_{load}, p_{bc}, p_{obs}, \dots)$$

where a is the leading finite-size scaling exponent reported in our manuscript, and F collects the influence of additional physical and observational controls. For example, p_{mat} may include elastic properties, frictional or damage parameters and characteristic strength; p_{geom} may describe aspect ratios, slope or fracture geometry; p_{load} and p_{bc} may describe loading rate, stress state and boundary conditions; and p_{obs} may include sampling interval, detection threshold, monitoring technique and data quality. More rigorously, these quantities would enter F through appropriate dimensionless combinations.

This formulation makes clear how we interpret the observed scaling. The power-law dependence V^a represents the dominant cross-system dependence associated with event size, whereas the function F accounts for secondary controls that shift individual events above or below this leading-order trend. In logarithmic form

$$log\, T = a\, log\, V + log\, F(\dots)$$

so that much of the observed scatter around the volume scaling can naturally be interpreted as variability in F. The remarkable empirical result is therefore not that all other parameters are irrelevant, but that, despite their variability across very different systems, a substantial fraction of the variation in precursory duration can be organized by a single leading dependence on event size.

This is analogous to finite-size scaling more generally: microscopic properties, geometry, constitutive behaviour and forcing may affect amplitudes, crossover scales and corrections to scaling, while a dominant size dependence remains identifiable over a sufficiently broad range. The present dataset spans multiple hazard classes and many orders of magnitude in event volume, which is precisely the regime in which such a leading-order dependence can emerge even though secondary controls remain important.

The purpose of the present analysis is to identify this broad cross-system scaling, rather than to construct a complete multivariate predictive model for T. We already report separate regressions by hazard type, providing a first assessment of whether the leading dependence is consistent across the principal event classes. In the revised manuscript, we will additionally examine residuals by hazard type to determine whether systematic offsets or trends remain among these classes, and we will make more explicit that the proposed scaling should be understood as a first-order relation with potentially important corrections contained in F.

A comprehensive multivariate or hierarchical decomposition of F is, however, not robustly supported by the present literature-derived dataset. Material properties, geometry, loading conditions, monitoring method, temporal resolution and related quantities are not reported consistently across studies; several are not defined in a directly comparable way across hazard types; and subdivision of the dataset along these dimensions rapidly produces groups that are too small to constrain additional effects reliably. Introducing many covariates under these conditions would risk replacing a robust leading-order result with an underconstrained model and substantial overfitting.

We will therefore clarify this limitation and frame the present result accordingly: T proportional to V^a should be regarded as the leading term of a more general relation rather than as the assertion that volume alone determines precursory duration. Quantifying the structure of F, and in particular identifying which material, geometrical, loading and observational variables account for the residual variability, is an important next step that will require larger datasets with more systematically reported metadata.

The interpretation in terms of finite-size criticality is plausible, but the data primarily show an empirical association between duration and size. The abstract currently moves quickly from an empirical association to a mechanistic interpretation in terms of system-spanning correlated deformation and universality. I think this should be toned down. The data may be consistent with finite-size scaling, but they do not uniquely demonstrate it.

Reply: We thank the referee for this comment. We agree with the general principle that an empirical scaling relation should be distinguished from its physical interpretation, and we will make this distinction explicit. We would, however, argue that the observed relation is not merely a generic empirical association between duration and size. The relation exhibits additional structure that carries physically meaningful information beyond the existence of a positive correlation alone.

The primary empirical result is a robust power law scaling between precursory duration T and failure volume V for mechanically driven geohazards,

$$T \propto V^{0.35 \pm 0.05}.$$

Because a characteristic system size scales as $L \sim V^{1/3}$, this empirical result corresponds to

$$T \propto L^{1.05 \pm 0.15}.$$

The significance of this result is that the exponent is, within uncertainty, close to unity. This is precisely the form expected from a dynamic finite-size scaling relation

$$T \sim L^z$$

with a dynamic exponent $z \simeq 1$. Thus, finite-size criticality is not invoked solely because duration happens to increase with event size; the measured exponent provides a quantitative connection between the observed scaling and the finite-size scaling framework.

There is also an important element of selectivity in the data. Landslides, glacier breakoffs, and rockbursts show broadly consistent scaling exponents, despite their very different materials, geometries, loading conditions, and characteristic sizes. Volcanic eruptions, by contrast, do not exhibit a significant dependence of precursory duration on erupted volume. This distinction is physically meaningful. The mechanically driven systems considered here can plausibly evolve through progressive interaction, damage, slip, and correlation growth within a finite deforming volume, whereas volcanic unrest involves additional processes such as magma transport, pressurisation, phase changes, and evolving rheology, for which erupted volume need not represent the relevant finite mechanical system size. The fact that the scaling emerges for the mechanically driven classes but not indiscriminately for all geohazards therefore argues against the interpretation that it is simply a generic consequence of applying the same analysis to heterogeneous datasets.

The finite-size interpretation also follows naturally from the dynamical framework developed in the manuscript. Near a critical instability, spatial correlations are expected to grow as interacting damage or slip processes that progressively organise. Their spatial extent may be represented by a correlation length ξ, with an associated temporal scale $\mathcal{T}$. Dynamic scaling gives

$$\mathcal{T} \sim \xi^z.$$

In an infinite system these characteristic scales may continue to grow as the critical point is approached, whereas in a real finite system the spatial correlation length is necessarily truncated by the characteristic system size, $\xi \sim L$. The corresponding temporal scale is therefore also cut off at

$$\mathcal{T}_{\max} \sim L^z.$$

Within this framework, the observed near-linear relation between T and L has a direct physical meaning: it is consistent with the time required for correlated deformation to organise from smaller scales up to the scale of the failing system.

We therefore agree that the present observations do not constitute a mathematically unique proof of finite-size criticality, and we will avoid wording that could be interpreted in this overly strong sense. However, we do not think that the mechanistic interpretation should be reduced to a merely speculative possibility. The combination of (i) a statistically robust scaling over more than ten orders of magnitude in V, (ii) an exponent quantitatively close to the finite-size dynamic-scaling expectation $z \simeq 1$, (iii) consistency across several distinct mechanically driven hazard classes, and (iv) the absence of the same scaling in volcanic eruptions provides specific and nontrivial support for the finite-size criticality interpretation.

We will therefore revise the manuscript to distinguish clearly between observation and interpretation while retaining this physical conclusion. Rather than stating that the scaling uniquely demonstrates finite-size criticality, we will state that the near-linear relationship between precursory duration and characteristic system size provides quantitative evidence consistent with finite-size critical organisation, in which the precursory duration reflects the time required for correlated deformation to develop up to the system scale. This formulation, in our view, accurately reflects both the strength of the empirical evidence and the appropriate limits of mechanistic inference.